ERNEST ORLANDO LAWRENCE

BERKELEY NATIONAL LABORATORY

International Review of Frameworks for Impact Evaluation of Appliance Standards, Labeling, and Incentives

Nan Zhou, John Romankiewicz, Edward Vine,
Nina Khanna, and David Fridley
China Energy Group
Environmental Energy Technologies Division
Lawrence Berkeley National Laboratory

December 2012

This work was supported by the China Sustainable Energy Program of the Energy Foundation and Collaborative Labeling and Appliance Standards Program through the U.S. Department of Energy under Contract No. DE-AC02-05CH11231.

Disclaimer

International Review of Frameworks for Impact Evaluation of Appliance Standards, Labeling, and Incentives

Nan Zhou, John Romankiewicz, Edward Vine, Nina Khanna, and David Fridley
China Energy Group
Environmental Energy Technologies Division
Lawrence Berkeley National Laboratory

Executive Summary

In recent years, the number of energy efficiency policies implemented has grown very rapidly as energy security and climate change have become top policy issues for many governments around the world. Within the sphere of energy efficiency policy, governments (federal and local), electric utilities, and other types of businesses and institutions are implementing a wide variety of programs to spread energy efficiency practices in industry, buildings, transport, and electricity. As programs proliferate, there is an administrative and business imperative to evaluate the savings and processes of these programs to ensure that program funds spent are indeed leading to a more energy-efficient economy. The field of energy efficiency program evaluation grew out of this imperative and has the following primary objectives:

1) to measure and verify the impacts of a specific energy efficiency program
2) to evaluate the processes of a specific energy efficiency program
3) to inform program managers' and policymakers' decision making in both assessing market potential and improving program design

Within these three objectives, there is an emphasis on both measuring the impacts of a program as well as analyzing the processes by which a program works. Because evaluation is becoming increasingly important, it is becoming a standard practice in certain regions to set aside a portion of a program's budget (for example, 2-4%) for evaluation, even before the project has begun. The categories and types of evaluations are summarized in Table 1. The most common types of evaluation are potential, process, and impact evaluations, although, recently, market assessment and market effects evaluation are being conducted more widely.

- Potential or feasibility evaluation – also known as *ex-ante* evaluation – is implemented during a program planning phase to predict the program impacts.
- Process evaluation is performed during program implementation to see, for example, if program participants are interacting with the program and receiving programs benefits as planned.
- Finally, impact evaluation – also known as *ex-post* evaluation – is used to determine what energy savings and environmental impacts a program produced. This paper will focus on both

ex-ante and *ex-post* evaluation methodologies within the context of a specific set of energy efficiency programs – appliance standards, labeling, and incentive programs.

Table 1. Categories and types of program evaluation

Evaluation category	Phase of implementation	Evaluation type	Assessment level
Formative	Pre-program planning phase	Market assessment (includes characterization, baseline)	Market, portfolio, program
		Potential or feasibility (sometimes known as *ex-ante*)	Portfolio, program, project
	Implementation phase – ongoing	Process	Portfolio, program
Summative	Implementation phase – ongoing and/or *ex-post*	Impact	Program, project, measure
		Market effects	Market, portfolio
		Cost effectiveness	Portfolio, program, project

Source: Adapted from Brown et al. 2007

The basic research question in impact evaluation is: What were the true effects produced by a program or intervention in terms of energy savings (as well as other impacts, such as changes in electricity demand and carbon emissions), separated out from what would have otherwise occurred absent the program or intervention? Energy efficiency evaluations calculate energy savings, which is energy that was not used. Trying to estimate such a counterfactual case is very tricky, and evaluators are constantly working to remove as much uncertainty from this process as possible. Data type, quality, and source relate directly to the cost of an evaluation as well as to the related uncertainty in the outputs of that evaluation. For example, evaluators often spend extra money to reduce uncertainty by conducting more field measurements or larger surveys. Figure 1 shows this general trend of smaller amounts of uncertainty with higher cost of evaluation and vice versa.

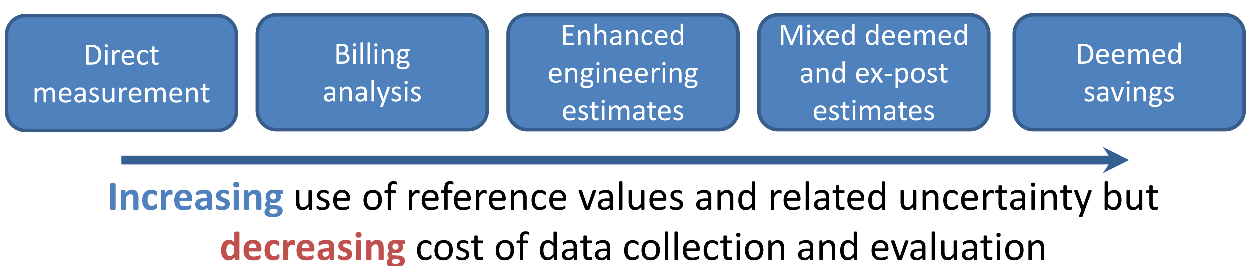


Figure 1. Energy use measurement and estimation methods in program evaluation

Source: Adapted from EU EMEES and ODYSSEE projects

This paper references over 60 evaluation studies from the U.S., E.U., Australia, and other countries, and it looks at 30 studies in depth for unique evaluation methodologies. Of those, ten studies are used more thoroughly throughout the paper to show examples of evaluation calculations and methodologies for appliance standards, labeling, and incentives.

A number of studies on standards evaluation have been done in the U.S. and Australia, and in particular, *ex-ante* evaluation of standards plays a large role in the U.S. standards development. Figure 2 outlines a framework for *ex-ante* and *ex-post* evaluation of appliance standards. While seven steps are outlined in the figure and explained in full detail in the report, the general methodology can be broken down into three main parts: 1) stock model, 2) baseline setting, and 3) *ex-post* evaluation options.

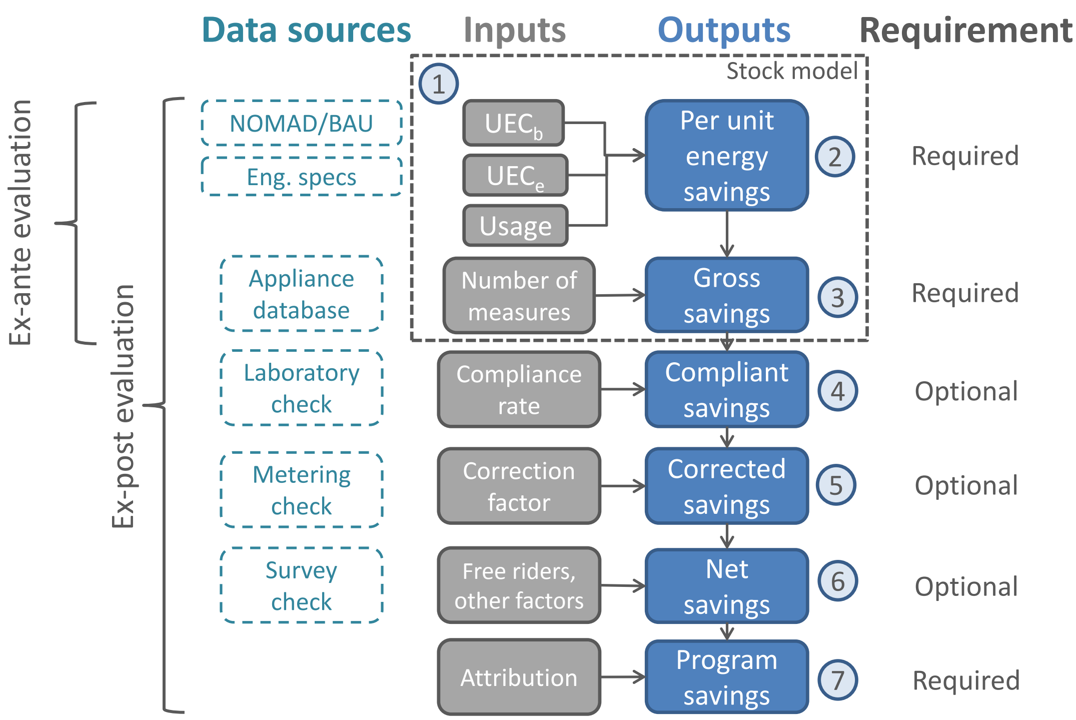


Figure 2. *Ex-ante* and *ex-post* evaluation frameworks for standards

A stock model is key for most standards and labeling evaluations; it keeps track of the efficiency breakdown and energy consumption of a fleet of appliances based on engineering specifications, lifetimes, and unit energy consumption. Baselines help define how the efficiency of the appliance fleet would have improved without the standard, often referred to as naturally occurring market adoption (NOMAD). The main options here are as follows:

1. Frozen baseline: the efficiency of new products remains constant in the base case
2. Improvement baseline: where historic unit energy consumption (UEC) data exist, the efficiency of new products improves at a similar rate of historic autonomous efficiency improvement, which declines into the future
3. Market share baseline: where data on market share for models of different efficiencies exist, a baseline efficiency can be estimated for future years
4. Bass model baseline: the most advanced curve fitting of market adoption of energy-efficient products to predict NOMAD

An example of an improvement baseline from a recent evaluation done in Australia on standards and labeling programs for refrigerators is shown in Figure 3.

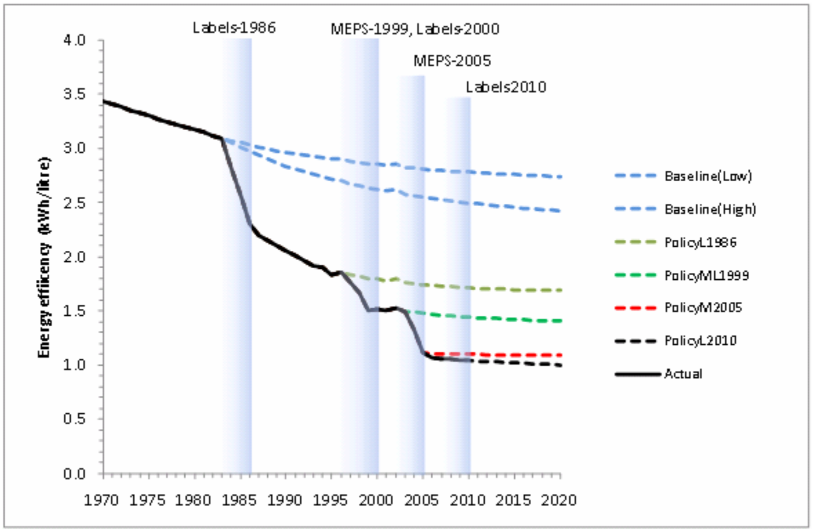


Figure 3: Average efficiency of group 5T refrigerators under different policy scenarios

Finally, *ex-post* corrections can be taken into account since manufacturer-claimed performance and laboratory performance are often very different from installed (field) performance, as shown in Figure 4. A compliance rate would technically be the performance differential between and a manufacturer's claim for a certain product and that product's laboratory claim when submitted for verification testing. A correction factor accounts for the difference between a manufacturer's claimed performance for a product and how that product actually performs when installed in the field. In these types of analyses, *ex-post* corrections are not common, as most standards evaluations simply rely on *ex-ante* estimates of the proposed program. However, correction factors are used in many of the U.S. Department of Energy's *ex-ante* evaluations of appliance standards during the development process.

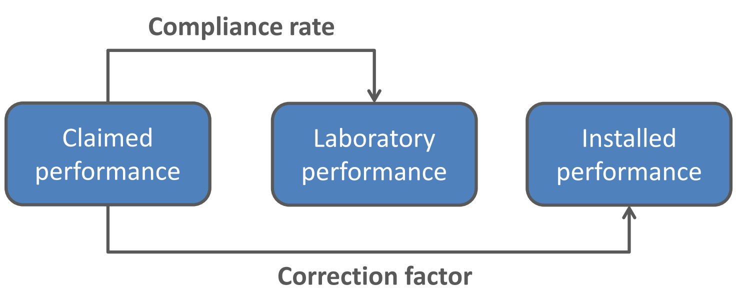


Figure 4. Relations of compliance rates and correction factors for performance data

The paper then outlines the required and optional data requirement for *ex-ante/ex-post* evaluation of standards. The stock model, comprised of current stock, UEC, market saturation, lifetime, and future shipments forecasts, is required for an *ex-ante* evaluation, while moving to an *ex-post* evaluation may require additional datasets such as usage adjustment factors and compliance rates (see Table 2). A stock model with as much data as possible on historic sales trends by efficiency and product class is the foundational starting point, however, for a successful evaluation.

Table 2. Required and optional data requirements and sources for ex-post/ex-ante analysis of standards

Data type	Used in *ex-ante* or *ex-post*	Required or optional	Data source
Annual energy consumption per unit (UEC)	*Ex-ante, ex-post*	Required	Manufacturer test data
Existing stock	*Ex-ante, ex-post*	Required	Market data, government statistics
Market saturation (ownership, market shares)	*Ex-ante, ex-post*	Required	Market surveys
Lifetime or retirement function	*Ex-ante, ex-post*	Required	Manufacturer test data
Future shipment forecasts	*Ex-ante*	Required	Historic market data, government forecasts
Usage adjustment factor (UAF)	*Ex-ante, ex-post*	Optional	Metered test data
Naturally occurring market adoption (NOMAD)	*Ex-ante, ex-post*	Optional	Historic market data
Compliance rate	*Ex-post*	Optional	Metered test data
Real shipments/sales	*Ex-post*	Optional	Market data
Site-to-source energy conversion factors	*Ex-ante, ex-post*	Optional	Power plant energy data
Emission factors	*Ex-ante, ex-post*	Optional	Power plant emission data

When performing an evaluation of a labeling program, as opposed to a standards program, the stock and shipments data (whether projected or real) will need to be broken down by efficiency category (see Figure 5) in order to calculate energy savings. The European Union (EU) has a long history of categorical labeling and has experience in the evaluation of categorical labeling programs.

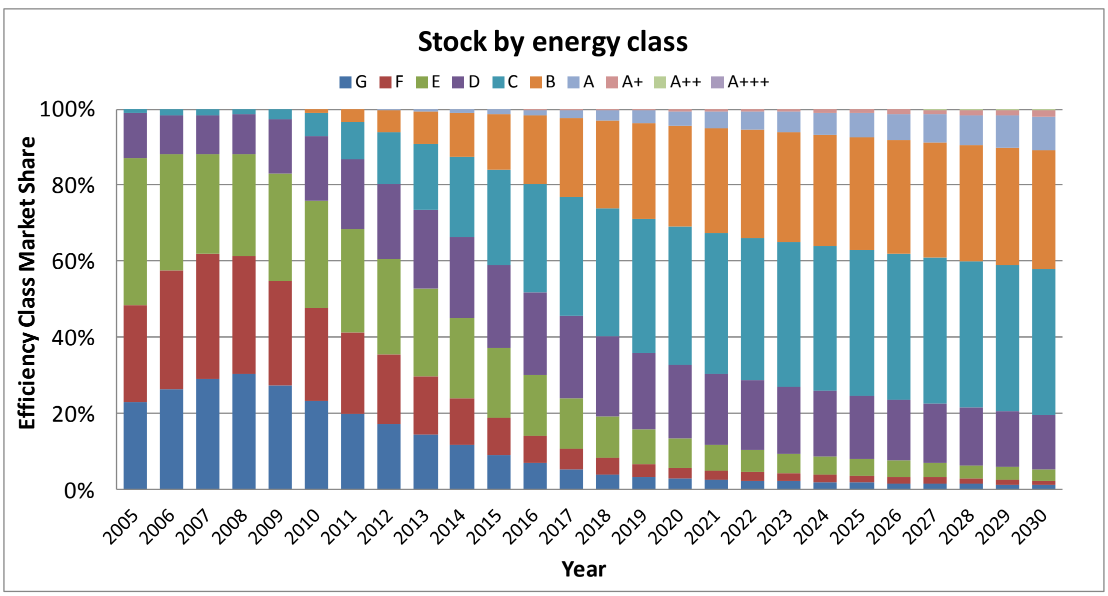


Figure 5. E.U. Estimation Tool Project Sales by Energy Class

Source: Larsen et al. 2012

Lastly, the evaluation of incentive programs largely revolves around the question of gross savings versus net savings. The gross savings amount can simply be described as the number of measures (for instance, number of rebates given out on a highly efficient refrigerator) multiplied by the expected energy savings per unit per year, which itself could be a deemed or measured value. Definitions of net savings vary from state to state in the U.S. depending on many factors. In some states, net savings “adjusts” the gross savings estimate by subtracting the savings from “free riders” (i.e., those who were going to take an energy efficiency action regardless of whether or not they received an incentive. For example, the California definition for net savings is:

$$CA: Net\ savings = Gross\ savings - free\ riders$$

In contrast to California, other states account for not only free ridership (which typically decreases the gross savings amount) but also participant spillover, whereby a participant may take more energy saving actions upon receiving a certain incentive which may have impacted his or her knowledge and decision making surrounding energy efficiency. As such, participant spillover tends to increase the gross savings amount. Free ridership and participant spillover are usually determined via telephone surveying of a selection of program participants. In terms of the actual quantification of the gross energy savings in the first place, Figure 6 shows the various evaluation methods from the prescriptive method of deemed savings (commonly seen for retail CFL –compact fluorescent lamp - rebates) to the custom method of monitoring or metering of whole building or HVAC efficiency measures. In practice, most states use a combination of these methods.

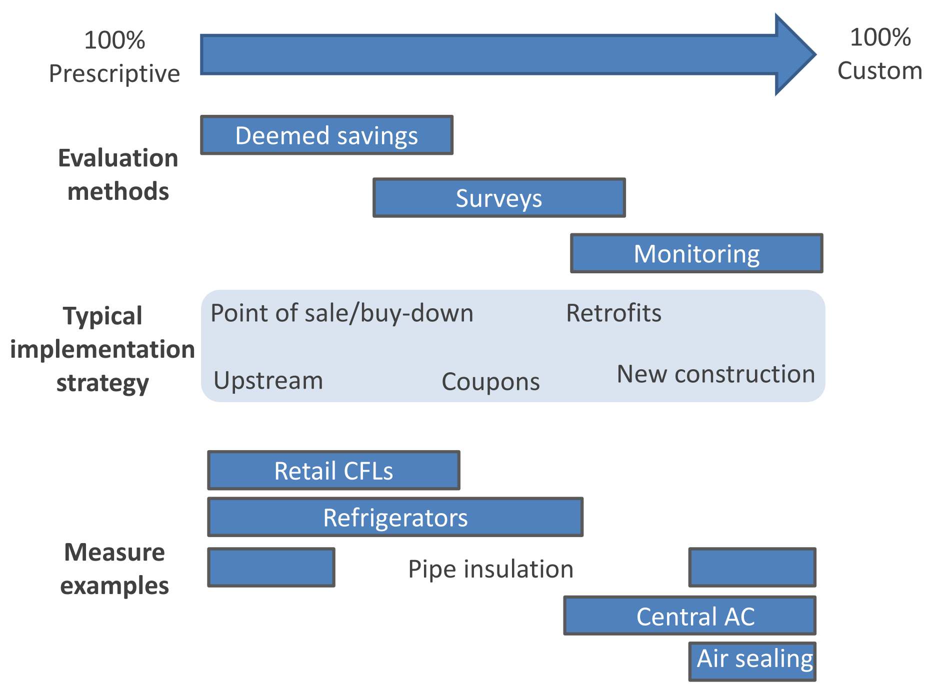


Figure 6. Prescriptive vs. custom evaluation methods and associated implementation strategies and measure examples

Source: Adapted from Dent and Enterline 2012

Table 3 outlines optional and required data points for the evaluation of a typical incentives policy and further breaks down the requirements into those for a gross energy savings calculation and those for a net energy savings calculation.

Table 3. Required and optional data requirements and sources for impact evaluation of incentive programs

Data type	Required or optional for gross energy savings	Required or optional for net energy savings	Data source
Annual energy savings per unit product or per building	Required	Required	Deemed values, IPMVP[1] approach, or statistical analysis
Number of participants and non-participants	Required	Required	Surveys
Normalizing factors (HDD, CDD)	Required	Required	Weather station
Free riders	Optional	Required	Surveys, econometric methods, deemed value
Participant spillover	Optional	Required	Surveys, econometric methods, deemed value
Market effects (participant & nonparticipant spillover)	Optional	Required	Surveys, econometric methods & market analysis
Site-to-source energy conversion factors	Optional	Optional	Power plant energy data
Emission factors	Optional	Optional	Power plant emission data

Most evaluation handbooks' first recommendation is to get good quality data and to start collecting it as early as possible. Indeed, data sources and collection methods are the basis of any evaluation, and that is why this paper has a section on data requirements for each of the evaluation topics covered – appliance standards, labeling, and incentives.

[1] International Performance Measurement and Verification Protocol (IPMVP)

Table of Content

List of Acronyms

ACEEE	American Council for an Energy Efficient Economy
BAU	Business as usual
BUENAS	Bottom-Up Energy Analysis Tool
CEC	California Energy Commission
CFL	Compact fluorescent light bulb
CLASP	Collaborative Labeling and Appliance Standards Program
CPUC	California Public Utilities Commission
DOE	U.S. Department of Energy
EMV	Evaluation, measurement, and verification
EPA	U.S. Environmental Protection Agency
EUF	Energy use factor
IEA	International Energy Agency
IEPEC	International Energy Program Evaluation Conference
IPMVP	International Performance Measurement and Verification Protocol
LBNL	Lawrence Berkeley National Laboratory
LEAP	Long-range Energy Alternatives Planning tool
MEPS	Minimum energy performance standard
NAPEE	National Action Plan for Energy Efficiency
NOMAD	Naturally Occurring Market ADoption
RECS	Residential Energy Consumption Survey
S&L	Standards and Labeling
SWEUF	Shipment weighted energy use factor
TRM	Technical Resource Manual
UAF	Usage adjustment factor
UEC	Unit energy consumption

1. Introduction

Standards remove inefficient products from the market (the cut-off volume in Figure 7), while labeling informs consumers and intermediaries (contractors, energy utilities, retailers). Both standards and labels stimulate technological innovation, as consumers become more informed and manufacturers compete. While standards will induce a relatively immediate effect, labeling policies take longer to induce market transformation. Market transformation describes how energy efficient technologies are adopted in the market, from initial adoption to full saturation. Full market transformation is achieved when barriers to the adoption of such technologies are reduced to the point where public support is no longer needed and consumers will adopt efficient technologies on their own. Rebates often encourage consumers to buy a more expensive, but more energy-efficient product than they would normally buy, inducing more market transformation and innovation on the part of manufacturers. These related processes are shown in Figure 7.

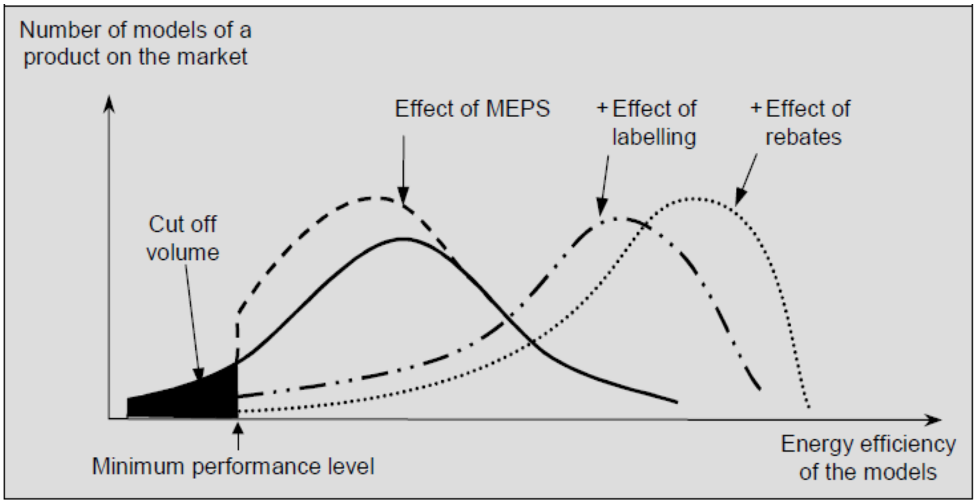


Figure 7. Illustrative distribution of efficiencies due to program impacts from MEPS, labeling, and rebates

Source: CLASP 2004

As standard, labeling, and incentive programs have expanded in the past three decades in the United States (U.S.), European Union (E.U.), Australia and other developed countries, evaluation techniques to quantify associated energy savings and market transformation have proliferated. As similar programs expand rapidly in China, India, Brazil, and other developing countries, there is a growing need to transfer techniques, methodologies, and experience from the collected experience.

The importance of program evaluation cannot be underemphasized. Not only can program evaluation quantify and verify the energy savings and environmental impacts of a specific energy efficiency program, but it can also help policymakers and program managers improve program design and better understand market potential. Additionally, program evaluation plays an important role in justifying government or utility spending on energy efficiency programs by showing that energy efficiency is a cost-effective resource.

The objective of this report is to review common evaluation methodologies and frameworks in the U.S., E.U., and Australia for appliance standards, labeling, and incentive programs, so that other countries can learn from these efforts and, if appropriate, apply the tools and methods to their situation. Section 2 provides an overview of program evaluation and its roles, types, and uncertainties. Section 3.1 reviews the collective experience in evaluation to date, providing a summary of key studies before leading into the specific methodologies. Sections 3.2, 3.3, and 3.4 review the methodologies for standards, labeling, and incentives evaluation, respectively, and also outline data requirements and sources for each evaluation type, as well as example calculations.

2. Overview of Program Evaluation

To understand program evaluation, we first need to understand the various types of energy efficiency programs that exist. Section 2.1 provides this overview, followed by Section 2.2 which explains the role of program evaluation. Section 2.3 summarizes the major types of evaluation, including impact, process, and cost-effectiveness evaluations. Finally, Section 2.4 discusses data requirements and related uncertainty for evaluation. As certain types of data collection can be relatively expensive, there will always be a correlation between the data quality and the cost of an evaluation.

2.1. Types of Energy-Eficiency Programs

Energy efficiency (EE) programs are varied and include everything from regulation-based measures (such as building codes or minimum energy performance standards [MEPS]) to informational campaigns (such as product labeling) to economic incentives (such as rebates and subsidies). EE programs typically have at least one of the following characteristics, as described by the National Action Plan for Energy Efficiency's Evaluation Guide (NAPEE 2007):

1. Resource acquisition
2. Market transformation
3. Codes and standards
4. Education and training

Resource acquisition refers to a program whose primary objective is to directly achieve energy and/or demand savings through specific actions or measures. These types of programs are very common in many U.S. states as well as E.U. member states. Energy savings generally take the form of physical energy savings (kWh of electricity, tons of oil or coal, cubic meters of gas) or of demand savings (kW of electricity). From energy savings, CO_2 savings can be calculated using a variety of methodologies and assumptions.

Some energy efficiency programs set market transformation as their goal; in one definition by the California Public Utilities Commission (CPUC), market transformation is defined as "long-lasting sustainable changes in the structure or functioning of a market achieved by reducing barriers to the adoption of energy efficiency measures to the point where further publicly-funded intervention is no longer appropriate in that specific market" (CPUC 2008). Market transformation typically describes technology development, from initial research and development, to marketplace introduction, to widespread adoption. While energy efficiency programs can help energy efficient technologies gain market share and increase adoption, codes and standards will ensure a minimum level of energy efficient technology is being used, whether in vehicles, appliances, buildings, or other energy end uses. Some utilities and energy efficiency program administrators are working on not only changing technology adoption but also the general population's attitudes, knowledge, and awareness (AKA), a growing area of research in market transformation (SoCal Edison 2009). Education and labeling are one of the primary avenues for increasing AKA.

Some policymakers, such as the CPUC, define the energy efficiency potential of any given program in light of total achievable potential, as shown in Figure 8. Technical potential can be defined as the "complete and instantaneous penetration of all energy efficiency measures" in a particular application that are technically feasible from an engineering perspective, without being constrained by economics or other barriers to getting these measures installed. Economic potential would pare the technical potential down to measures where the incremental costs of each measure are less than the savings or cost of other resources (such as new power plants), without being constrained by other barriers. Achievable potential would take into account the amount of funding that a given program has and other policy and administrative barriers. Finally, the naturally-occurring potential refers to the amount of savings that would occur absent the program or other market interventions. In California, naturally occurring energy savings are subtracted from the program savings because the California Energy Commission (CEC) incorporates naturally occurring savings into their energy demand forecasts, and the CPUC's mandate is to provide estimates of additional savings above the baseline. Other countries and U.S. states follow similar methodologies (as will be discussed in detail in this report). With good program design, the program potential can in theory reach the maximum achievable potential. In practice, program costs estimates for achieving the maximum potential are often underestimated and other barriers affect installations, so that it is often not possible to meet that maximum potential. It should be possible over time, however, to come very close to maximum achievable, with good program design and a portfolio of policies that accounts for these barriers and that is continuously improved.

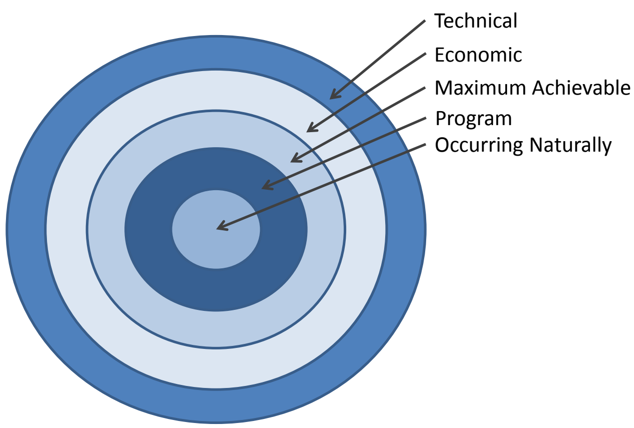


Figure 8. The potential of energy efficiency

Source: Ting and Rufo, 2010

2.2. The Role of Energy-Efficiency Program Evaluation

In recent years, the number of energy efficiency policies has grown very rapidly as energy security and climate change have become top policy issues for many governments around the world. The need to evaluate those policies has grown just as quickly. Energy efficiency program evaluation has the following primary objectives:

1) to measure and verify the impacts of a specific energy efficiency program;
2) to evaluate the processes of a specific energy efficiency program;

3) to inform program managers' and policymakers' decisions in both assessing market potential and improving program design.

The field of evaluation is often referred to in much of the literature as "evaluation, measurement, and verification" or EM&V. Here, the "M&V" refers to measuring or monitoring energy usage and verifying installations and assumptions (e.g., hours of use) from specific energy efficiency measures, projects or programs while the "E" refers to evaluating the energy savings from specific measures, projects or programs. In this context, programs represent a group of projects that have similar technology characteristics and applications, such as a utility program that provides rebates for residential clothes washers (Vine 2012, NAPEE 2007).

Typically, evaluations are commissioned by either governments or electric utilities. In both cases, outside contracting firms are often used to perform the evaluations, although in some cases governments perform their own evaluations in house. Governments and utilities not only use program evaluation to quantify and verify the energy savings and environmental impacts of a specific energy efficiency program, but also to help them improve program design and to better understand market potential. Additionally, program evaluation plays an important role in justifying government or utility spending on energy efficiency programs by showing that energy efficiency is a cost-effective resource. For this reason, when the budget for an energy efficiency program is being prepared, money is often set aside ahead of time for program evaluation during and after program implementation. This budget set aside is considered a best practice in the field of evaluation. In California, typically 4% of the program budget is set aside for evaluation in the planning phase.

2.3. Program Evaluation Types

As stated in Section 2.2, program evaluation can both assess the impacts (such as energy savings) or the processes of any given program, in order to better assess program design in the past and inform better program design in the future. Within these functions and goals set forth in the previous section, program evaluation can be divided up into six types as shown in Table 4.

Table 4. Categories and types of program evaluation

Evaluation category	Phase of implementation	Evaluation type	Assessment level
Formative	Pre-program planning phase	Market assessment (includes characterization, baseline)	Market, portfolio, program
		Potential or feasibility (sometimes known as *ex-ante*)	Portfolio, program, project
	Implementation phase - ongoing	Process	Portfolio, program
Summative	Implementation phase – ongoing and/or *ex-post*	Impact	Program, project, measure
		Market effects	Market, portfolio
		Cost effectiveness	Portfolio, program, project

Source: Adapted from Brown et al. 2007

Evaluations that are *formative* in nature take place either prior to program implementation (planning phase) or during program implementation, and focus on market assessment and program implementation processes. Evaluations that are *summative* in nature focus on impacts (e.g., energy savings and market impacts) and are completed during program implementation or shortly after program implementation (also known as *ex-post* evaluation).

As a program or a portfolio is planned, energy efficiency practitioners will likely engage in some type of market assessments in order to characterize energy efficiency practices in whatever field or technology they are looking at (e.g., retailers selling energy-efficient refrigerators). When designing the program, they will also undertake a feasibility evaluation to assess the potential energy savings of the measures and projects that they are proposing. This type of evaluation is known as *ex-ante* evaluation. Market assessments and *ex-ante* evaluations of energy savings are key elements of analysis during the development of energy efficiency standards.

While a program is being implemented, process evaluation is often used to examine the interaction between the program and its participants (e.g., consumers, homeowners, and commercial building owners) to see if there is room for improvement in the program design and implementation. For instance, if an energy efficiency labeling program is implemented, but most consumers misread or do not understand the information on the label, then this is information that the program administrator will want to know as soon as possible, so that appropriate changes can be made to the label, the delivery process, or some other aspect of the program. If an energy efficiency product rebate program is implemented, the program administrator may want to understand how long it takes a given consumer to redeem a rebate, as this may influence consumer perception and willingness to participate in the rebate program.

Evaluations that quantify impacts, market effects, and cost effectiveness are the most common assessments included under summative evaluation. Impact evaluations usually look at changes in energy use and demand, but it is not uncommon for non-energy benefits such as CO_2 emissions reduction, health benefits, job creation, and water savings to also be evaluated. For evaluating market transformation, a market effects evaluation will quantify changes in the energy efficiency marketplace and the uptake or adoption rate of various technologies.

Finally, for assessing cost-effectiveness, evaluations often conduct a cost-benefit analysis. The evaluation will include the results of the impact evaluation: typically, energy savings and the cost of program implementation (e.g., program administrator costs, customer costs, etc.). Typical costs for a standards or labeling program, for instance, may be incurred in the development of the standards and associated test procedures, administration of any related training or educational campaigns, and associated enforcement. If there is an incentive program, then the total money spent on subsidies, rebates, tax exemptions, or other types of financial benefits are also calculated. Costs and savings accrued for the consumer over the lifetime of a given installed technology also need to be considered. Costs could include any initial upfront investment in the energy efficiency measure as well as related repair and operation and maintenance (O&M) costs. Operation costs, the product of energy price and energy usage, will be compared between the energy efficient technology and a baseline technology that is less energy efficient, resulting in a net stream of energy and monetary savings.

Commonly used cost-effectiveness evaluation methods in the U.S. include:

- Total resource cost (TRC) test
- Program administrator cost (PAC) test
- Participant test
- Societal test (NAPEE 2007)

The total resource cost test (TRC) will include both the program administrator) costs and savings and the participant (consumer) costs and savings. A societal cost test is similar to a program administrator cost (PAC) test, but will also take into account externalities, such as air pollution reduction or other environmental improvement, that benefits society as a whole. A recent study by the American Council for an Energy Efficient Economy (ACEEE) looked at 45 different state energy efficiency programs and their evaluation methods. In this survey, 71% of the states consider the TRC to be their primary test, 15% indicate the Societal Test, and 12% used the Program Administrator test.

Of the evaluation types highlighted in Table 4, the primary focus of this report is the methodologies and tools for impact evaluation, although process evaluation will also be briefly discussed, as it has been a very important component of many labeling programs around the world. In addition, the best evaluation practice is to conduct both process and impact evaluations of the same program, since one can learn from the other (Kushler et al. 2012).

2.4. Basic Research Question of Impact Evaluation

The basic research question of impact evaluation is: What were the true effects produced by a program or intervention in terms of energy savings, separated out from what would have otherwise occurred absent the program or intervention? As described in a recent paper on key evaluation issues (Vine et al. 2012), the two most common issues faced in answering this question are: (1) definition of net savings, and (2) technical issues with measurement.

Gross energy savings is equal to the actual energy consumption minus the baseline energy consumption. For instance, consider the case of a customer who was using an incandescent light bulb but then switched to a compact fluorescent light (CFL) bulb because of a special promotional rebate offered at a local retail store. In this case, the gross savings per year would be equal to the difference in those two bulbs' energy consumption multiplied by the hours of use per year. Net energy savings adjusts the estimate of gross savings by accounting for three factors:

- <u>Free riders</u>: those participants who would have taken an energy saving action without the program/ incentive offered
- <u>Participant spillover</u>: program participants who as a result of the program took additional energy saving actions beyond those incentivized
- <u>Market effects</u>: any savings that occurred due to program influence on the market (also called non-participant spillover)

Generally, when accounting for these factors, free riders will reduce gross savings, while participant spillover and market effects will increase gross savings. Within the U.S., definitions of net savings vary from state to state (Vine et al. 2012). On the one hand, California's ratepayer funded energy efficiency programs define net savings as the gross savings of the program minus a free rider fraction. On the other hand, New York's energy efficiency programs define net savings as the gross savings minus the free rider fraction plus participant spillover and market effects. See the equations below.

$$CA: Net\ savings = Gross\ savings - free\ riders$$

$$NY: Net\ savings = Gross\ savings - free\ riders + market\ effects + participant\ spillover$$

In the recent ACEEE survey mentioned previously, the study found that 26% of states report gross savings in their evaluation results, 53% of states report net savings, and the remaining 21% use both gross and net savings, sometimes for different purposes (Kushler et al. 2012).

Beyond the issues of defining net savings, there are also technical issues with measurement that program evaluators must face. The basic issue is that program evaluators must document how the program changed behaviors or transformed the market without knowing what would have occurred otherwise (the "counterfactual"). In other words, the identification of baselines from which gross savings are estimated is very tricky. Additionally, the wide array of public policies and market inventions on energy efficiency that may be applicable in any given geography are both numerous and complex. This makes it difficult to sort out the net effects of any single program. For instance, Minimum Energy Performance Standards (MEPS) for appliances are often paired with labeling now, making it difficult to evaluate the distinct impact of each policy.

The result of this uncertainty has been a diversification of methodologies into both bottom-up and top-down categories. While bottom-up evaluation relies on surveys and quantification (of participants and specific measures) to inform inputs into engineering-based models, top-down evaluation uses aggregate consumption indicators (such as total energy consumption or per-capita energy consumption) as inputs to macro-economic demand models. There is an increasing interest in top-down evaluation as efficiency goals and the number of interacting policies increase. The drawback is that it is often hard to determine changes in energy demand due to specific energy efficiency programs versus other macroeconomic factors such as energy prices, income and indirect rebound effects, and structural changes. Many scholars contend that while bottom-up methods are more likely to overestimate energy savings (due to free riders and direct rebound effects), top-down methods tend to underestimate them. For evaluation of a portfolio of programs, some evaluators are beginning to implement both bottom-up and top-down methods concurrently (Ecofys et al. 2006).

2.5. Data and Uncertainty in Program Evaluation

Primary and secondary data quality is a major focus for any program evaluation effort. The Collaborative Labeling and Appliance Standards Project (CLASP) outlined typical evaluation data types and sources in their standard and labeling guidebook (Table 5). For instance, in evaluating standards and labeling (S&L) programs, a national appliance database that represents the appliances currently being sold in the market (and their efficiency and price characteristics) is an important element used in most impact evaluations.

Table 5. Evaluation data types and sources

Data type	Main data sources
Customer and retailer awareness, knowledge, and decision making	Surveys/interviews of customers and retailers
Availability of products	Sales data from manufacturers, trade associations; surveys of manufacturers and retailers
Prices for efficient products	Surveys of customers, retailers, and manufacturers
Market penetration	Sales data from manufacturers, trade associations; surveys of program participants and non-participants
Energy use	Manufacturer data; independent laboratory test data; metered end-use data; engineering specifications; deemed savings
GHG emissions	Reported emissions factors; utility dispatch model data

Source: Adapted from Wiel and McMahon 2005

Energy efficiency evaluations calculate energy savings, which is energy that was not used. As noted above, trying to estimate such a counterfactual case is very tricky, and evaluators are constantly working to remove as much uncertainty from this process as possible. Data type, quality, and source relate directly to the cost of an evaluation as well as to the related uncertainty in the outputs of that evaluation. For example, evaluators often spend extra money to reduce uncertainty by doing more field measurements or larger surveys.

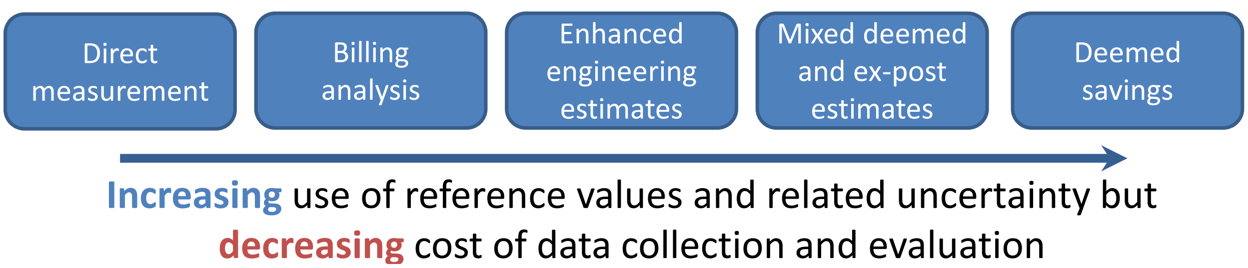


Figure 9. Energy use measurement and estimation methods in program evaluation

Source: Adapted from E.U. EMEES and ODYSSEE projects

If one was just to look at gathering energy use data described in Table 4, many different estimation or measurement methods could be used from direct measurement to deemed savings (see Figure 9). Direct measurement (e.g., using energy meters) could be used to get end use load data from equipment or appliances. This type of measurement can be costly when accounting for materials and labor, although costs may reduce in the future as smart appliances with wireless communication abilities develop. Billing analysis can also be used, if the evaluators are able to get the cooperation of the utility company. Although in the case of appliances, changes in energy use from one appliance may be too small to monitor given that one household (on one bill) uses many different appliances and lighting, and the change in energy use from that particular appliance may be difficult to discern in the utility bill.

In contrast to these types of measurement are estimation techniques, such as deemed savings. Many evaluation organizations in the U.S. have developed technical reference manuals (TRMs) that list standard savings values with an associated certainty. For example, a program that incentivizes the purchase of a CFL through a rebate may use a "deemed savings" value based on previous similar evaluation studies that estimate the savings that a CFL will produce over its lifetime in comparison to a baseline incandescent light bulb. Estimation and use of deemed savings values may be accurate enough for some evaluation studies and for some energy efficiency measures (particularly, measures that have a long history and whose energy savings are relatively predictable, based on a few key assumptions (e.g., refrigerators and CFLs)).

The tradeoff between evaluation cost and evaluation uncertainty will be a theme throughout the next section as evaluation methodologies and related data requirements are described. Evaluations rely on data, and the use of primary vs. secondary sources will have a big impact on costs and uncertainty depending on various conditions.

3. Overview of International Impact Evaluation Methodologies and Tools for Appliance Standards, Labeling, and Incentives

Program evaluation methodologies and tools have been in development for a number of decades now, with evaluation of energy efficiency programs dating back to the 1970s in the U.S. As policies to mitigate climate change increase in number, there has been more and more interaction and exchange on program evaluation methodologies and initiatives to share such methodologies with developing countries. The International Energy Program Evaluators' Conference (www.IEPEC.org) has been running since 1983 and is one example of such exchanges.

This section will first provide a quick overview of major evaluation studies performed to date on appliance standards, labeling, and incentives (outlined in Table 6 below). Then, Sections 3.2 to 3.4 will outline the main methodologies used in those studies along with related data requirements and example calculations.

3.1. Introduction to International Practices in Impact Evaluation

Figure 10 outlines a general evaluation methodology for standards, labeling, and incentives. Each input will require data and/or assumptions, and this section will highlight the various data sources, data gathering methods, and assumptions used in evaluation. Steps 1 through 7 are described on the following page to provide a brief overview of terminology and methodology used throughout section 3.

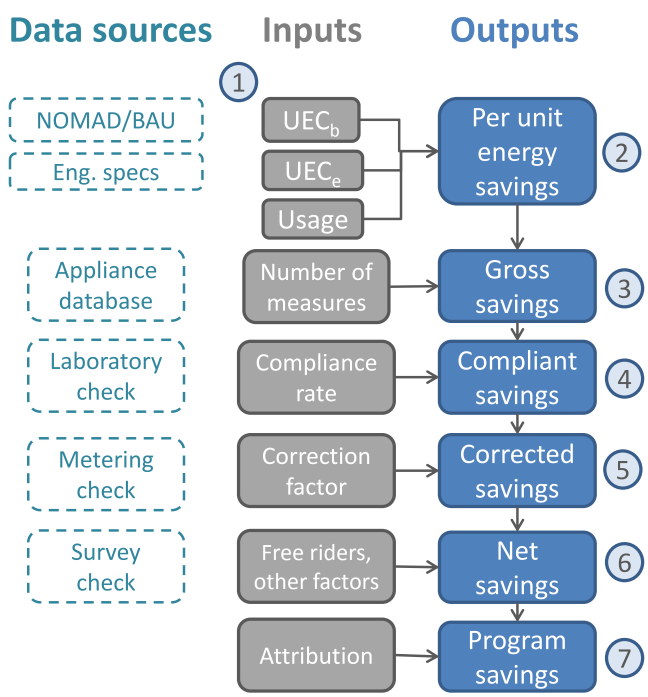


Figure 10. General evaluation methodology for standards, labeling, and incentives, including data sources, inputs, and outputs

Setting the Baseline

Setting the baseline against which energy savings are measured is one of the most important steps in program evaluation. Since standards, labeling, and incentive programs are all seeking to push more efficient appliances and products into the market, the questions are: what is the efficiency of the products that they are replacing or displacing and what would have been purchased if the program had not been put in place? What would manufacturers have produced and retailers stocked on shelves without appliance standards? What type of refrigerator would a consumer have purchased without a rebate for a more energy efficient model? A **business as usual (BAU)** case or **naturally occurring market adoption rate (NOMAD)** is set to determine several key evaluation elements, including the **baseline unit energy consumption (UEC_b)** for these counterfactual cases. Examples for how baselines are set in the case of standards and labeling (S&L) are provided in Sections 3.2 and 3.3, respectively.

Calculating Per Unit Energy Savings

The **unit energy consumption (UEC_e)** for the unit replacing or displacing the baseline unit can be gathered from engineering specification documents from the manufacturer. Subtracting UEC_e (energy use of the energy-efficient product) from UEC_b (energy use of the product in the baseline) under the correct usage conditions will give the per unit energy savings. The usage of the old and new products may vary in certain cases (e.g., by climate zone).

Calculating Gross Savings

Unit energy savings can be multiplied by the **number of measures** or units sold (e.g., products that received an energy efficiency rebate or products sold under a new efficiency standard) to calculate **gross savings**. A typical equation for a refrigerator would look like this.

$$Gross\ savings = number\ of\ refrigerators * (UEC_b - UEC_e) * usage * (1 + IF_{HVAC})$$

where UEC_b is the energy use (kWh/h) of the refrigerator in the baseline case, UEC_e is the energy use of the rebated refrigerator, usage is the number of hours of use, and IF_{HVAC} is the interaction factor for measuring an heating, ventilation, and cooling (HVAC) system on a refrigerator's operation. For the purposes of streamlining these calculations for large numbers of products and product groups, spreadsheet models are usually developed to track product vintage, efficiency, hours of use/lifetime, price, and other factors. Steps 1 through 3 are the only steps for some evaluations that focus on gross energy savings. The corrections that take place in steps 4 through 7 are the focus of net energy savings evaluations. The first three steps are also commonly seen in *ex-ante* program evaluations or as part of an impact analysis (such as in Australia or the U.S.) for proposed standards under development.

Correcting for Non-compliant Products

Frequently, appliances use more energy than claimed by the manufacturer for compliance with a standard or labeling requirement. If this is the case, then any non-compliant products will reduce the overall energy savings of an appliance standards, labeling, or incentive program. Potentially, a program administrator could do verification testing of a representative sample size at third-party laboratories in order to quantify the **compliance rate** or the difference between claimed energy efficiency and actual energy efficiency. In fact, reliable data on compliance rates are not that widely available in many places around the world, so many program evaluations often skip this step. Additional explanations will be provided in the proceeding sections.

Correcting for Performance Variance

On the other hand, even if an appliance is tested in a laboratory and it is proven to perform at its claimed energy efficiency, it still may operate differently in someone's home. For instance, a refrigerator will operate differently under field conditions than it performs in a laboratory test chamber with a constant temperature. Here, **correction factors** can be used based on metering test studies done in the field. Metering is a relatively expensive data gathering method, so it is only performed for evaluations with sufficient budget. There is potential for overlap between correction factors and compliance rates. For instance, in gathering energy use data on an appliance, a meter can test whether the product is compliant relative to rated performance as well as whether the product has any performance variance due to installation or environmental circumstances.

Additional Baseline Adjustments to Arrive at Net Savings

Baseline adjustments can be made either at the beginning of the evaluation calculation or towards the end. For S&L evaluations, the baseline is typically set depicting what would have happened in the absence of those programs as described in step 1. For incentive evaluations, it is common to introduce some additional adjustments for **free riders** and **participant spillover** as described in Section 2.3. In the case of free riders, the original baseline will account for the product being replaced or displaced, while the free rider proportion will account for those participants who would have gone ahead with purchasing a more efficient product even without the incentive that was offered. Participant spillover will account for additional energy saving actions that the participants may have made due to the incentive program. **Market effects** adjustments can be made for incentive programs as well as S&L programs, since they try to account for any larger level of long-term market transformation towards energy efficiency that has occurred. Information for these types of calculations on free riders, participant spillover, and market effects is usually gathered using various surveying techniques that will be described in detail in Section 3.4.

Attributing Savings to Various Programs

In the last step, a specific savings amount has to be attributed to a specific program. This is important when there are various programs that overlap in their influence on one sector's energy efficiency. It is common for labeling and incentive programs to have some overlap, for instance, since both programs are pushing the consumer to purchase an above average efficiency product. Like free riders in step 6, **attribution factors** are also calculated using some mixture of surveying and expert input.

Table 6. International studies and methodologies/protocols reviewed in standards, labeling, and incentives evaluation

	Standards	Labeling	Incentives
U.S.-Federal MEPS	Meyers et al. 2008, 2011 McNeil et al. 2012 DOE 2011a, 2011b		
U.S.-ENERGY STAR		EPA 2011 (process/impact) Homan et al. 2011 (impact)	
California	TecMarket Works 2006 Quantec 2007		TecMarket Works 2004, 2006 Itron DEER database
Other U.S. states			Vine et al. 1999 Vine et al. 2012 Kushler et al. 2012 NAPEE 2007 Vermont 2010
Australia	EE Strategies 2002, 2010		
E.U.		Larsen 2012 Larsonneur 2009 Luttmer et al. 2006 Vreuls 2005 Waide 1997	Broc et al. 2009 SRC et al. 2001
Canada	NRCan 2012		
International	Vreuls 2005 Agra Monenco 2000a, 2000b		IPMVP 2002, updated 2011 Vreuls 2005

3.1.1. United States – Federal level

Lawrence Berkeley National Laboratory (LBNL) has conducted a number of national impact assessments for standards for major residential and commercial appliances, using dynamic BAUs where assumptions were made on baseline efficiency improvements over time in the absence of new standards. The data for making these assumptions came from a mixture of sources including trade associations (e.g., the Association of Home Appliance Manufacturers (AHAM)) and U.S. Department of Energy (DOE) Technical Support Documentation (TSD) for related appliance standards development (Meyers et al. 2008, Meyers et al. 2011, DOE 2011a, DOE 2012b). LBNL is also working on developing better methodologies for dynamic baseline setting based on Bass curves in support of its Bottom-Up Energy Analysis System (BUENAS) model (McNeil et al. 2012).[2] Figure 11 shows an example for data points for the ENERGY STAR dishwasher market share for 2000-2008 that fits a general Bass curve shape (Van Buskirk 2012). Bass curves are widely used to describe market adoption, and there is some evidence that these types of curves should also have applicability to NOMAD, in the absence of any standards or labeling programs. Often, it is difficult to gather data for products in the absence of S&L programs, however, since federal standards for most major product categories have existed since the late 1970s and since the mid-1990s for labeling programs.

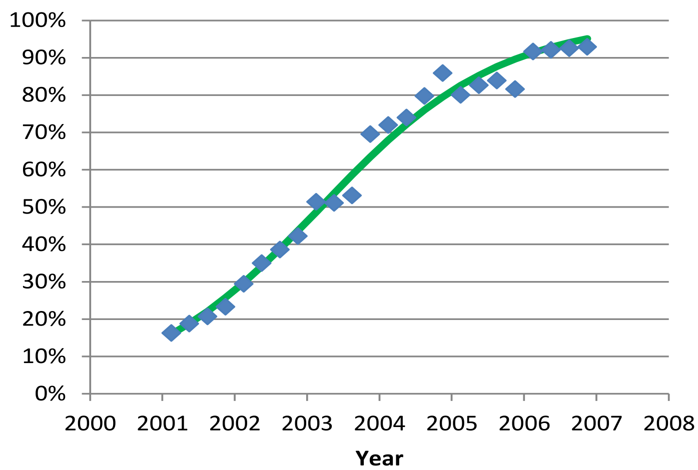


Figure 11. ENERGY STAR market share for dishwashers, 2000-2008

Source: Van Buskirk 2012

While LBNL's evaluation studies have incorporated dynamic baselines where possible, estimates for compliance rates and correction factors have yet to be incorporated. Section 3.2.1 discusses in more depth some of the calculation techniques used in these evaluation studies for dynamic baselines, compliance rates, and correction factors.

Also at the federal level, the U.S. Environmental Protection Agency (EPA) carries out impact evaluations for its certification labeling system ENERGY STAR. Key indicators such as ENERGY STAR awareness (see Figure 12) and ENERGY STAR impact on purchasing decisions are quantified through annual surveys (EPA

[2] The Bass diffusion model describes the process of how new products get adopted as an interaction between users and potential users and is often used in product forecasting.

2011). LBNL has also conducted impact studies on the resultant energy savings from this labeling program (Homan et al. 2011), and the methodologies are discussed in Section 3.3.1.

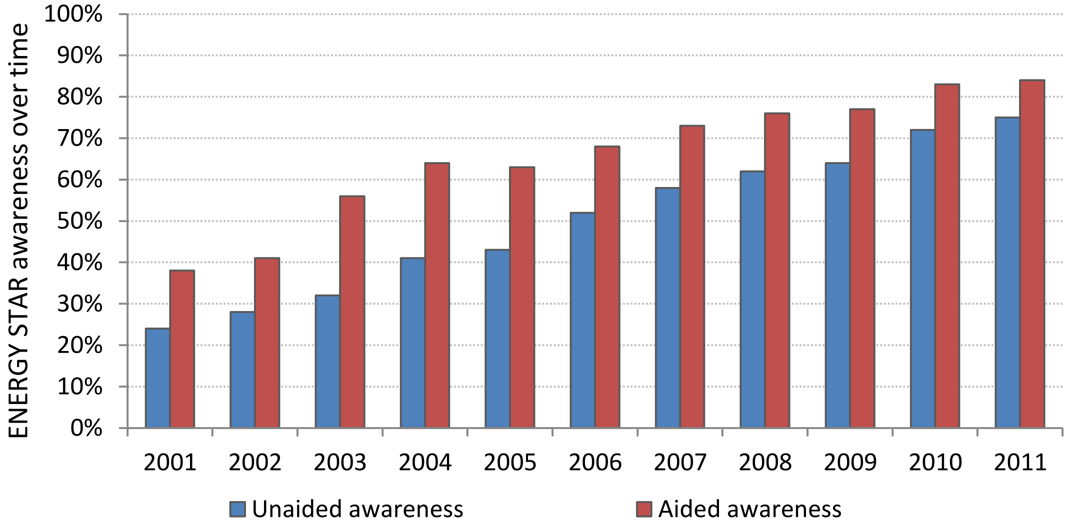


Figure 12. ENERGY STAR awareness over time

Note: aided vs. unaided awareness refers to whether or not the ENERGY STAR label was shown to the survey participant upon questioning

Source: EPA 2011

3.1.2. California

California is a leader in energy efficiency, and as such, has developed protocols and procedures for impact and process evaluation (TecMarket Works 2004, CPUC 2006, Vine et al. 2006). Particularly in the area of incentive programs for building, appliance, and lighting efficiency, these protocols have been very helpful for program evaluators. Alongside the development of protocols, the evaluation community in California developed the Database for Energy Efficient Resources (DEER) that contains "deemed savings" values for efficiency measures that are commonly installed in the marketplace. Replacement of incandescent bulbs with CFL bulbs is a common example. Section 3.4 goes into more depth on the use of this type of database (and other similar resources called Technical Resource Manuals or TRMs).

California has also established codes and standards in product areas that are not covered by federal standards. As such, they recently performed an impact evaluation of several state standards that they had implemented, which incorporated estimates for NOMAD and compliance rates, which will be reviewed in Section 3.2. California lists the following outputs for the evaluation of codes or standards:

- A listing of the technologies or practices influenced by the program that experienced an energy efficient code or standard change.
- A listing of the code and standard changes that will be addressed in the evaluation. (Items 1 and 2 may be the same, but also may be different if the evaluation is addressing a subset of the changes.)
- An estimate of the influence of the program on the code and standard changes for each technology or practice included in the evaluation.

- An estimate of the NOMAD rates for each technology or practice included in the evaluation.
- An estimate of the date when each code or standard change would have occurred without the program for each technology or behavior included in the evaluation.
- An estimate of the level of non-compliance expected for the technologies and practices covered in the evaluation over the period of time that savings are projected.
- An estimate of gross and net market-level energy impacts for the program as a whole and for each technology and practice covered in the program and for each utility territory funding the program. This estimate of impacts should not exceed a 30-year effects lifetime. (CPUC 2006)

3.1.3. Other U.S. States

There have been a number of studies by experts at LBNL, ACEEE, and the EPA looking at the wide range of evaluation practices found across the U.S. as mentioned earlier in Section 2.3 (Vine et al. 2012, NAPEE 2007, Kushler et al. 2012). Other states have developed their own regional or state based TRM's, similar to California's DEER database (e.g., VEIC 2010). Amidst the diversity of existing methodologies given the U.S.'s diverse regulatory space for electricity and energy efficiency, there have been efforts to consolidate methodologies and potentially make a national impact evaluation protocol for various energy efficiency measures. The State and Local Energy Efficiency Action Network (SEE) wrote a scoping study to identify issues with developing such a protocol (Schiller et al. 2011), while the National Renewable Energy Laboratory (NREL) is running the Uniform Methods Project, which has developed some draft methodologies for specific efficiency measures in residential lighting, commercial HVAC, and other areas. Relevant methodologies will be referenced in the following sections.

3.1.4. Australia

Australia has had a number of efforts in the evaluation of appliance programs. In 2002, Australia created its first draft framework for the evaluation of Australian Energy Efficiency S&L program. An additional program audit for labeling was performed in 2004/2005, with more comprehensive program audits for air conditioners and refrigerators performed as recently as 2010 (EE Strategies 2002, 2010). These studies are very helpful for their use of improving baselines and correction factors (see Section 3.2.1), as well as their strong data collection techniques (see Section 3.2.2).

3.1.5. European Union

European evaluation efforts have been headlined by two recent large-scale projects in support of E.U. energy efficiency targets. The first project was the Evaluation and Monitoring for the E.U. Directive on Energy End-Use Efficiency and Energy Services (EMEEES), which focused on bottom-up methodologies for 20 different energy efficiency measures (Larsonneur et al. 2009). The second project was the European Energy Efficiency Database (ODYSSEE), which focused on top-down indicators for energy efficiency as a route to monitor energy efficiency trends and policy measures in Europe (Lapillonne et al. 2009).

The EMEEES methodologies covered all forms of common programs including regulation (MEPS), labeling and training measures, financial incentives, and voluntary agreements. Section 3.3.1 reviews the methodologies relevant to labeling impacts on appliances, including refrigerators and freezers. In addition to creating bottom-up methodologies for 20 different measures, the EMEEES project also proposed a template for member states to establish National Energy Efficiency Action Plans (NEEAPs) and a methodology for the E.U. Commission to perform *ex-ante* assessments and *ex-post* evaluations of the NEEAPs.

The summary report for the EU's EMEEES project argued that top down methodologies are accurate enough for appliances and vehicles, if there were well defined statistical indicators of average specific energy consumption per unit, while bottom up methodologies were difficult due to free rider and multiplier effects which can be very costly to quantify. If the measure saved more than 40 million kWh in annual electricity savings, more than 100 million kWh in annual energy savings, or greater than 5% of any individual member state's Energy Services Directive target, bottom up methodologies were justified. If a member state had only limited data about the measure, they used EU default values. If a member state had very measure-specific data, then there were specific harmonized guidelines they could follow to calculate the savings (Wuppertal Institute 2009).

In addition to reviewing the EMEEES bottom-up methodology for appliances, other studies were referenced given the E.U.'s long history of categorical labeling programs (Larsen 2012, Waide 1997). Figure 13 shows the impact that categorical labeling can have on refrigerators. Section 3.3.1 will go into more detail on how this market share data can be used to calculate energy impacts.

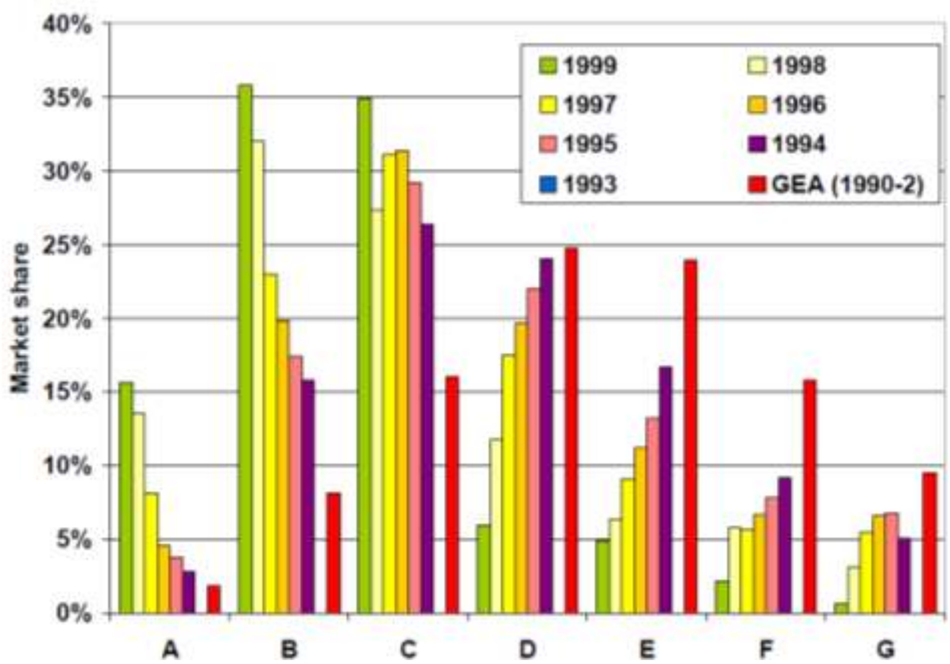


Figure 13. Impact of E.U. labeling + MEPS scheme on refrigerator market share, showing increasing market share of higher efficiencies (A, B, C) over time,

Source: Weil 2002

3.1.6. Other Countries and Studies

One of the projects in the International Energy Agency's (IEA) Demand Side Management Programme performed an overview study on energy efficient program evaluation in 2005 and looked at general frameworks for measuring and verifying impacts as well as evaluating effectiveness and efficiency of policy measures (Vreuls 2005). Specific to impacts from standards, labeling, and incentive programs, Table 7 shows that there is overlap between the inputs, outcomes, and impacts as defined by the IEA. All of the programs require inputs of funding (albeit at different levels) as well as human resources. Each one can lead to an increase in the sales of products of higher energy efficiency as well as increased awareness of energy efficient products (and energy efficiency, in general). All of the programs, if successful, will lead to some quantification of energy savings (and associated CO_2 savings). In the long-term, market transformation will occur leading to sustainable changes in the energy efficiency marketplace. The principal difference is in the output of each program. The implementation of a standard is very different from the implementation of an incentive program.

Table 7. Inputs, outputs, outcomes, and impacts of standards, labeling, and incentive programs

<table>
<tr><th></th><th>Inputs</th><th>Outputs</th><th>Outcomes</th><th>Impacts</th></tr>
<tr><td>MEPS</td><td rowspan="3">Funding (from industry and government) and human resources (program administrators and evaluators)</td><td>Tools for enforcing minimum standard</td><td rowspan="3">Increased awareness of energy efficient products, increase in sales of energy efficient products</td><td rowspan="3">Energy savings (and associated CO2 savings), market transformation (longer term)</td></tr>
<tr><td>Labeling/ education</td><td>Number of impressions, website hits, etc.</td></tr>
<tr><td>Incentives</td><td>Subsidized energy efficient products</td></tr>
</table>

Source: Vreuls 2005

A MEPS program will have a number of outputs including the related legislation, test protocol, and information material on MEPS. The outcomes will be designs and manufactured appliances that are compliant with the MEPS requirements, from which the impact will be energy savings. The IEA study outlines a number of indicators for both MEPS and labeling programs, along with three tiers of evaluation complexity depending on data evaluation and budget: comprehensive, targeted, and review.

Natural Resources Canada (NRCan) has performed impact evaluations on the MEPS for appliances in Canada. The data used in their studies are based on a biannual survey that NRCan's Office of Energy Efficiency conducts. The Canadian Appliance Manufacturing Association conducts confidential data surveys for each manufacturer to submit shipment data on their sales of major appliances. The study assumes a frozen baseline for calculating energy savings resulting from MEPS (Vreuls 2005).

In Thailand, the Electricity Generating Authority of Thailand (EGAT) ran a voluntary categorical labeling program for refrigerators and air conditioners. The EGAT's Office of Demand-side Management commissioned a process evaluation to gather qualitative data about behavior and attitudes of consumers and manufacturers, a market evaluation to assess the impact of the program on manufacturer decisions and market penetration, and an impact evaluation to assess the program's effect in terms of energy and demand savings. Surveys and interviews were conducted with 50 manufacturing and distribution firms as well as 2,000 households. Direct metering of air conditioner and refrigerator units combined with the data from the surveys and program data on product efficiency were used together to estimate energy and demand savings (Agra Monenco 2000a, 2000b).

The Cooperative Labeling and Appliance Standards Program (CLASP) researchers conducted a study on China's awareness and knowledge about the labeling program, similar to the ENERGY STAR awareness survey mentioned previously (Zeng et al. 2011). The study spanned 15 cities and counties in 10 provinces and municipalities and found that – of thousands surveyed – around 61.5% have some knowledge of the energy labeling system, with the majority becoming familiar with the label through in-store experiences and seeing the label on the displayed products themselves. The researchers hope to establish a replicable study to be able to continuously analyze and improve China's labeling system and related awareness. On the topic of standards evaluation, China now releases an annual white paper on its standards and labeling policies, which provides ex-ante evaluation estimates of the energy savings from new MEPS (CNIS 2012).

Finally, it is worth mentioning the International Performance Measurement and Verification Protocol (IPMVP). Monitoring and verification (M&V), a key component in the evaluation of energy efficiency projects, has been an evolving art and science since the late 1970's, when it was performed on ad-hoc, case-by-case basis with no available standards. Since that time, numerous M&V guidelines have been promulgated, but IPMVP is the most well-known M&V. The IPMVP was developed with U.S. DOE sponsorship and is currently managed by a non-profit organization (Efficiency Valuation Organization, EVO) that is developing new M&V material for publication as publicly available documents.

North America's energy service companies have adopted the IPMVP as the industry standard approach to measurement and verification. States ranging from Texas to New York now require the use of the IPMVP for state-level energy efficiency retrofits. The U.S. Federal Government, through the DOE's Federal Energy Management Program (FEMP), uses the IPMVP approach for energy retrofits in Federal buildings. Finally, countries ranging from Brazil to the Ukraine have adopted the IPMVP, and the Protocol has been translated into Bulgarian, Chinese, Czech, Hungarian, Polish, Portuguese, Russian, Spanish, Ukrainian and other languages.

The IPMVP provides a framework and definitions that can help practitioners develop M&V plans for their projects to verify energy savings. The IPMVP includes guidance on best practice for determining savings from efficiency, water conservation, and renewable energy projects. Typical M&V activities fall into four areas:

1. Prepare a site-specific M&V Plan
2. Define the pre-installation conditions that influence the baseline energy consumption
3. Define post-installation conditions that influence post-installation energy consumption
4. Conduct M&V activities to verify operation and achieved energy savings

To quantify energy savings, one or more of the following M&V techniques may be used: inspections, engineering methods, metering, statistical analyses, and computer simulation of system performance. Often, M&V involves the integration of several of these techniques. The IPMVP was built around a common structure of four M&V options. The purpose of providing several M&V options was to allow the user flexibility in the cost and method of assessing savings. A particular option is chosen based on the expectations for risk and risk sharing between the buyer and seller as well as project specific features. The options differ in their approach to the level and duration of the verification measurements. None of the options are necessarily more expensive or more accurate than the others. Each has advantages and disadvantages based on site specific factors and the needs and expectations of the customer. Project evaluators are expected to use one of these options for reporting on measured energy savings.

3.2. Impact Evaluation of Appliance Standard Programs

Ex-ante evaluation of appliance standard programs plays a large role in a number of countries' standards development process, whereby the impact on national energy demand can be estimated for different levels of proposed standards – essentially different UEC_e levels against the same baseline. Shipment projections are used to estimate the "number of measures", the number of products that will be sold under the new standard. Baselines can be set to estimate some initial level of market penetration for high efficiency products in the absence of the standard. These elements compose what is typically called an engineering-based model or a stock model in the literature. *Ex-post* evaluation can use the same estimates as *ex-ante* evaluation or it can update them based on data rather than projections, for example, using actual shipments or sales as opposed to shipment projections. Additionally, *ex-post* evaluation has the option of using a number of correction factors to get a more accurate estimate of total energy saved from the standard, as shown in Figure 14. This section will review methodologies for 1) stock models, 2) setting baselines, and 3) *ex-post* evaluation options. It will also review associated data requirements and provide calculation examples.

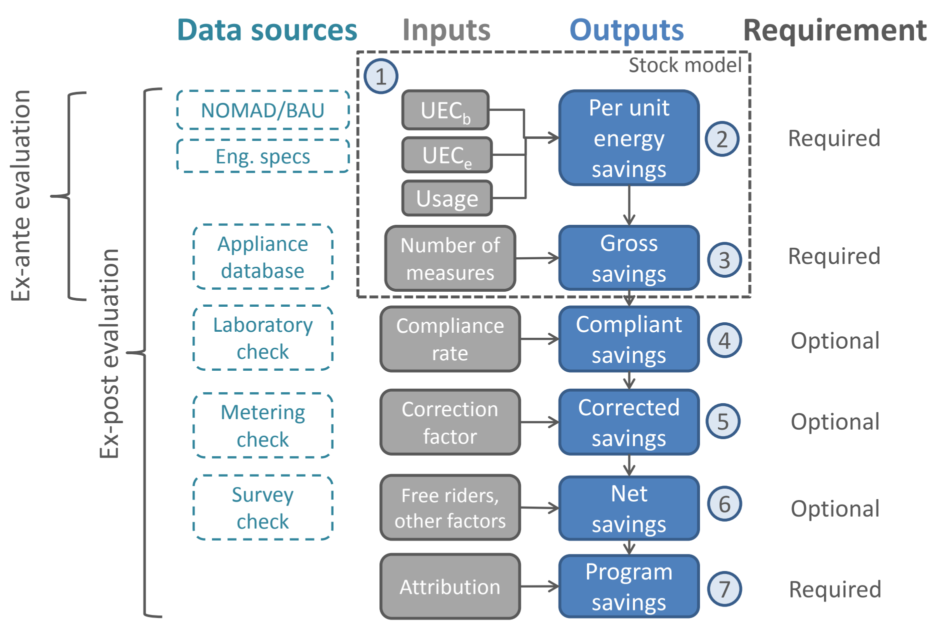


Figure 14. *Ex-ante* and *ex-post* evaluation frameworks for standards

3.2.1. Review of Existing Evaluation Methodologies

3.2.1.1. Stock Models

Stock model methodologies do not vary widely for appliances. Similar methodologies were found in evaluation or standards development studies performed in the U.S., E.U., and Australia. Typically, the model is simply an interaction between the existing stock, sales, and retirement of appliances, and their related energy consumption over time. The BUENAS methodology developed at LBNL (McNeil et al. 2012) uses the following equations for that interaction. First, there is an equation for total energy consumption in a BAU case.

(Eq. 1)
$$E_{BAU} = \sum_{age} Sales(y - age) * UEC_{BAU}(y - age) * Surv(age)$$

- *Sales (y)* = unit sales (shipments) in year y
- *UEC(y)* = unit energy consumption of units sold in year y
- *Surv(age)* = probability of surviving to *age* years

Unit sales (shipments) can be derived from increases in stock and replacements, if shipment data are not available.

(Eq. 2)
$$Sales(y) = Stock(y) - Stock(y - 1) + \sum_{age} Ret(age) * Sales(y - age)$$

- Stock(y) = Number of units in operation in year y
- Ret(age) = Probability that a unit will retire (and be replaced) at a certain age

The survival function and the retirement function are related by:

(Eq. 3)
$$Surv(age) = 1 - \sum_{age} Ret(age)$$

When sales and stock data are both not available, as is typical in many developing countries, these numbers are estimated by the number of households as well as by a diffusion equation, which is based on household incomes, urbanization, and electrification. These data types are more typically available.

Figure 15 shows the interaction of these functions in calculating gross energy savings.

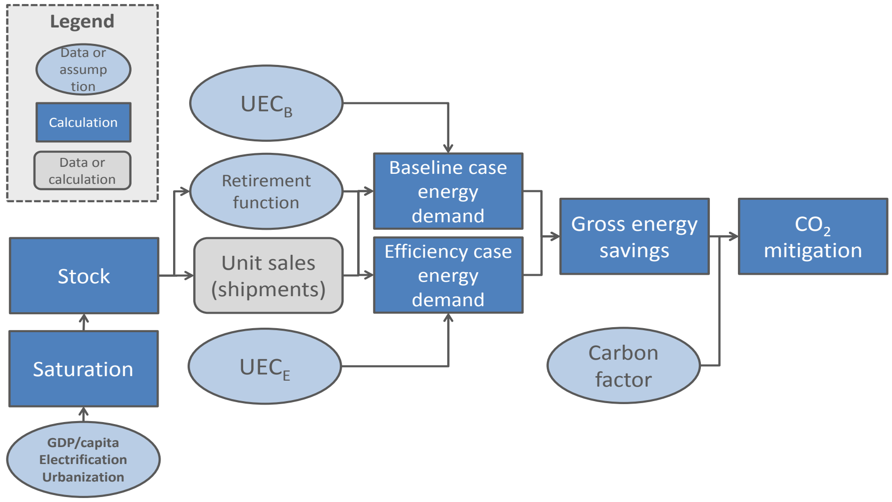


Figure 15. BUENAS methodology for *ex-ante* evaluation of gross energy savings

Source: McNeil et al. 2012

In their process for standards development in the U.S., the DOE calculates annual national energy savings (NES) between a baseline case (BASE) and an efficiency standards case (STD), as such:

(Eq. 4)
$$\mathrm{NES_y} = \mathrm{AEC_{BASE}} - \mathrm{AEC_{STD}}$$

The annual energy consumption (AEC) is calculated by multiplying the stock (or number of appliances) in each product class by its UEC according to the following equation and variables.

(Eq. 5)
$$AEC_y = \sum STOCK_V * UEC_V$$

- *AEC = national annual energy consumption, summed over vintages of stock for different product classes, STOCKV*
- *NES = annual national energy savings*
- *STOCKV = stock of product of vintage V that survive in the year for which DOE calculated annual energy consumption*
- *UECV = annual energy consumption per product in kilowatt-hours (kWh)*
- *V = year in which the product was purchased as a new unit*
- *y = year in the forecast*

Establishing UEC for baseline and standards cases is explained in the next section on baselines. It is interesting to note that the U.S. methodology also later incorporates a quantity known as a site-to-source conversion factor to account for losses associated with the generation, transmission, and distribution of electricity. Electricity savings are calculated as site energy (kWh), then converted to primary energy using the site-to-source conversion factor.

The stock equations are as follows:

(Eq. 6) $$Stock\ (j, age = 1) = Ship(j - 1)$$

(Eq. 7) $$Stock(j + 1, age + 1) = Stock(j, age) * [1 - prob_{Rtr}(age)]$$

- *Stock (j, age) = number of units of a particular age*
- *j = year for which the stock is being estimated*
- *Ship (j) = number of units purchased in year j*
- *ProbRtr(age) = retirement probability function*

Equation 6 states that the number of units that are one year old is equal to the number of new units purchased in the previous year, while Equation 7 is used to remove a fraction of the stock based on the retirement probability function. These retired appliances will then have to be replaced, which can be estimated using the following equation, where N is the year in which the model begins its stock accounting:

(Eq. 8) $$Rpl_p(j) = Stock(j - 1) - \sum_{age=0}^{ageMax} \sum_{j=N}^{j-1} Ship_j * prob_{Rtr}(age)$$

Finally, the shipments can be estimated with the following equation:

(Eq. 9) $$Ship(j) = Rpl(j) + NH(j) + Conv(j)$$

Where Ship is the total shipments in year j, Rpl is the replacement shipments in year j, NH is the shipment to new homes, and Conv is the shipment due to additional refrigerator purchases (second refrigerators, or an existing household's first refrigerator purchase). A survival function is also a key component in stock and replacement calculations. It will describe the probability that an appliance is still in use at a certain age according to the following equation, based on a Weibull distribution, a common equation for describing failure rates.

(Eq. 10) $$P(x) = e^{-\left(\frac{x-\theta}{\alpha}\right)^{\beta}}$$

- *P(x) = probability that the appliance is still in use at age x,*
- *x = appliance age,*
- *α = scale parameter, which would be the decay length in an exponential distribution,*
- *β = shape parameter, which determines the way in which the failure rate changes through time,*
- *ϑ = delay parameter, which allows for a delay before any failures occur.*

3.2.1.2. Setting Baselines

Setting baselines for standards evaluation is becoming more difficult. There is no longer much data on autonomous efficiency improvements over time in many countries, as S&L programs have increased in number. While the proliferation of these programs is good for energy efficiency, it does make setting baselines somewhat of a guessing game, although methodologies are improving. In general, standards evaluations have a number of baseline setting techniques to choose from:

1. Frozen baseline: the efficiency of new products remains constant in the base case
2. Improvement baseline: where historic UEC data exist, the efficiency of new products improves at a similar rate of historic autonomous efficiency improvement, which declines into the future
3. Market share baseline: where data on market share for models of different efficiencies exist, a baseline efficiency can be estimated for future years
4. Bass model baseline: the most advanced curve fitting of market adoption of energy efficient products to predict NOMAD

Frozen baselines are typically used when there is a lack of data on market-driven improvements, as is typically seen in many developing countries. The BUENAS model often employs frozen baselines for *ex-ante* evaluation of potential efficiency standards in countries like Brazil and South Africa, while the China National Institute of Standardization (CNIS) also uses frozen baselines in its annual white papers that have projections on the potential savings from new MEPS on various products (CNIS 2012).

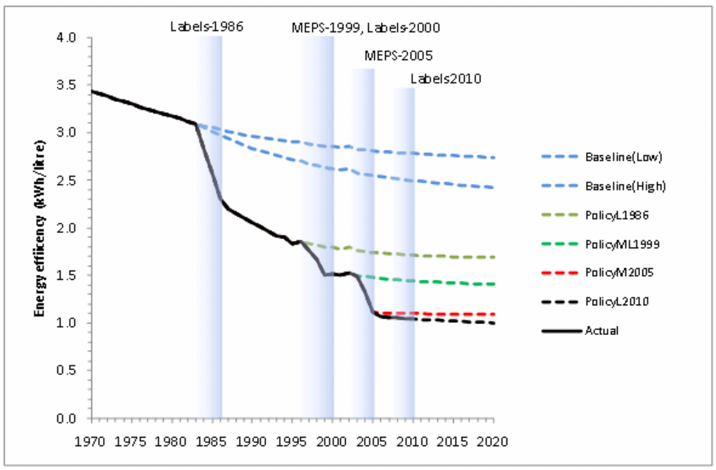


Figure 16. Average efficiency of group 5T refrigerators under different policy scenarios

Improvement baselines were used in the case of a recent, comprehensive evaluation of S&L programs for refrigerators in Australia (EE Strategies 2010). As shown in Figure 16, various S&L programs have had distinct impacts on the efficiency of refrigerators sold, including the first mandatory labeling implemented in 1986, updated labeling and MEPS implemented in 1999, and updated MEPS implemented in 2005. The evaluators in Australia had high quality shipments data (see the following section) and, therefore, they were able to chart an historic trend of weighted efficiency for all refrigerators sold in Australia for each product group. Each time a new policy was implemented, the previous rate of improvement (which had already slowed or flattened out from the previous policy) was extrapolated into the future as the new baseline for savings. From this baseline, cumulative electricity savings can be calculated by multiplying the average UEC by usage (24 hours a day, 365 days a year) by total units shipped, as shown in Figure 17.

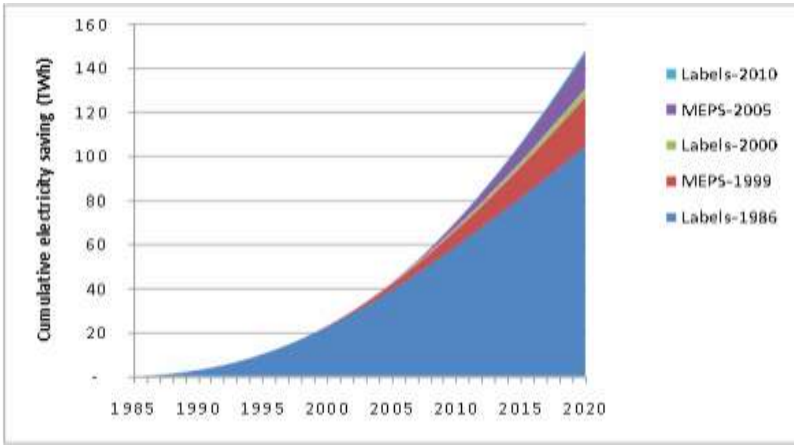


Figure 17. Cumulative electricity saved from S&L policies as compared to baseline

Source: EE Strategies 2010

Market share baselines are used in the *ex-ante* evaluation of potential standards levels for U.S. federal MEPS. Given that ENERGY STAR now plays a significant role in the evolution of appliance energy efficiency, it must be taken into account when forming a baseline. Typically, DOE will work with EPA to come up with estimates of where the ENERGY STAR market share of a particular appliance is heading in the absence of standards. Potential standards levels are set at 10%, 15%, 20%, etc. (with matching energy use factors [EUF] of 0.90, 0.85, 0.80, etc.) above a baseline and then market shares are estimated within these tranches. In order to get the UEC for baseline and standards cases, first the shipment weighted energy use factor (SWEUF) must be calculated according to the equation below.

(Eq. 11)
$$SWEUF_N = \sum_{N}^{6} EUF_N * Market\ share_N$$

Using the market share and EUF data found in Table 8, the SWEUF can be calculated and then multiplied by the baseline energy use for the particular category, which in this case is 539 kWh + 5*AV, where AV is the adjusted volume. The end result is that the base case takes into account some NOMAD of energy efficient products (in this case due to ENERGY STAR). Whereas in the case of frozen efficiency the SWEUF would be 1, in this case it is already 0.965 indicating that level of natural improvement. Similar market share assumptions are gathered for all product categories and then calculated against EUF to end up with UEC_b and UEC_e for use in *ex-ante* evaluation of any particular standard level.

Table 8. U.S. federal MEPS national impact analysis for standards development: Standard size top mount refrigerator-freezer efficiency market share, shipment weighted energy use factor, and average energy use for standards cases in 2014

Efficiency level (% less than base case)	Energy Use Factor (EUF)	Market share %						
		Base Case	Standard at efficiency level:					
			1	2	3	4	5	6
Base	1.00	78.2	–	–	–	–	–	–
1 (10)	0.90	4.2	82.3	–	–	–	–	–
2 (15)	0.85	9.4	9.4	91.7	–	–	–	–
3 (20)	0.80	8.3	8.3	8.3	100.0	–	–	–

4 (25)	0.75	0.0	0.0	0.0	0.0	100.0	–	–
5 (30)	0.70	0.0	0.0	0.0	0.0	0.0	100.0	–
6 (36)	0.64	0.0	0.0	0.0	0.0	0.0	0.0	100.0
Shipment weighted energy use factor (SWEUF)		**0.965**	**0.887**	**0.846**	**0.800**	**0.750**	**0.700**	**0.640**
Average energy use (kWh)		**520**	**478**	**456**	**431**	**404**	**377**	**347**

Source: DOE 2011a

The final example would be for baselines based on Bass-curves. Some markets, such as the E.U., have a wealth of data on appliance efficiency market share, having implemented labeling programs since the early 1990s. Figure 18 shows the market share for different efficiency grades of refrigerators in the E.U.

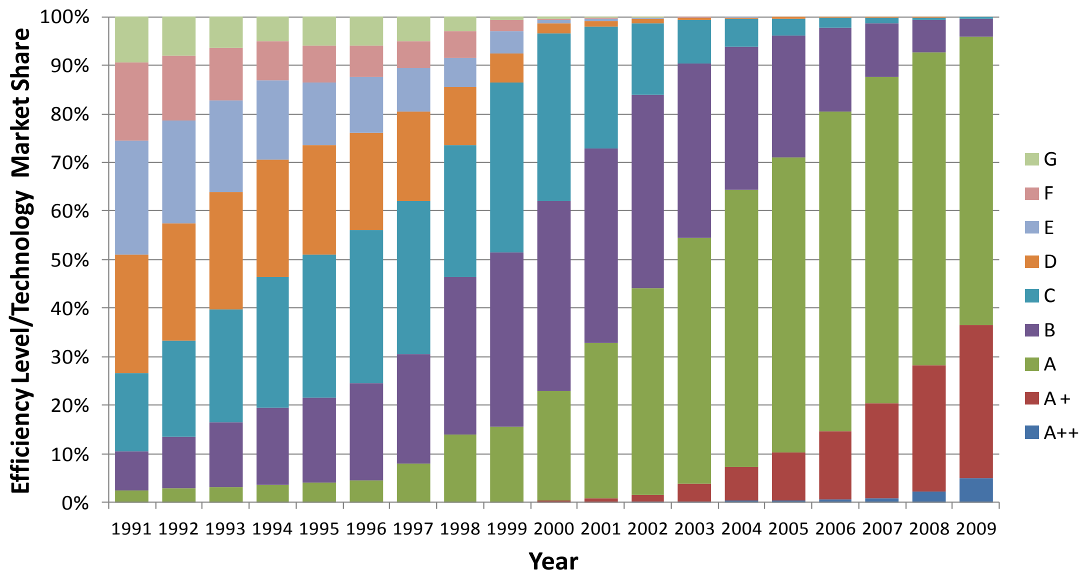


Figure 18. Evolution of market share for different efficiency grades for refrigerators in the E.U.

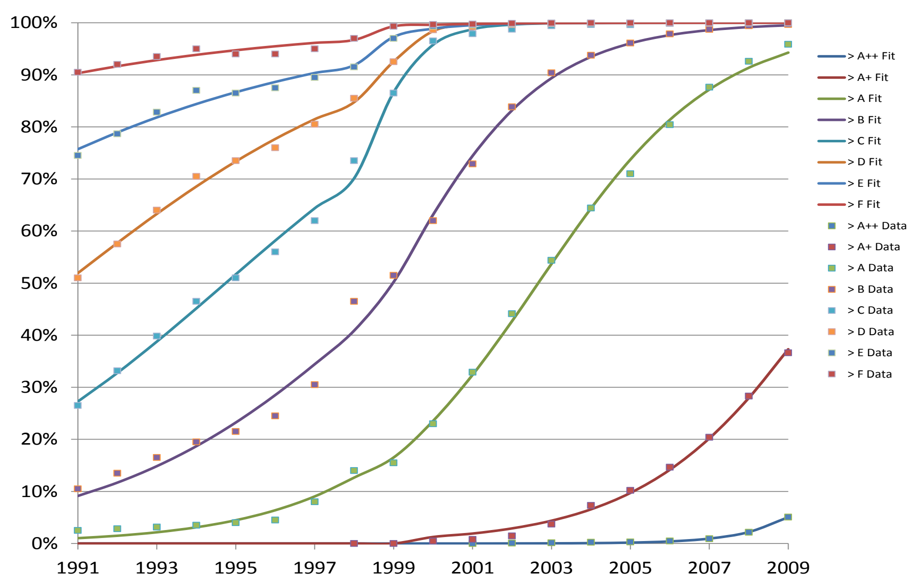


Figure 19. Mapped market share data points from Figure 18 with bass curve fits representing a combination of pre-MEPS and post-MEPS fits

Source: Van Buskirk 2012

Using these data, Bass curve fits can be developed that express how the market share of different grades evolved over time, using two different fits for pre-1999 and post-1999 time periods since a MEPS policy was introduced at that time (previously, only categorical labeling was in place) (Van Buskirk 2012). The research found that annual average improvement in the energy efficiency index of refrigerators was around 3.0% per year before 1999 and 6.6% after 1999, demonstrating the substantial impact that the MEPS policy played.

Table 9. Initial market penetration and NOMAD for select appliances and measures under MEPS in California

Measures and appliances	Market entry year	Initial market penetration 2006, %	NOMAD	
			2015	2030
Hardwired lighting, new residential	2000	8	35	56
Lighting controls under skylights, new	2000	7	38	51
Duct improvement, residential existing	1990	10	17	19
Ducts, nonresidential existing	1990	2	7	19
Consumer electronics-TVs	2000	41	76	83
Consumer electronics-DVDs	2000	24	58	61
Consumer electronics-audio players	2000	26	46	50
Res pool pumps, 2-speed, Tier II	1995	6	23	33
Pulse start metal halides	1992	26	46	57
General service incandescents	1970	47	50	52
Commercial dishwasher spray valves	1985	25	41	51
Unit heaters/duct furnaces	1965	50	58	65

Source: Quantec 2007

California has often developed standards at the state level for appliances and other building efficiency measures for which federal standards did not exist. In 2007, a study was performed on market adoption and compliance rates for these statewide codes and standards. Estimates were made on initial market penetration and NOMAD, as seen in Table 9. These estimates were made based on surveying experts in appliances and energy efficiency. They were surveyed using online simulation software to estimate a number of parameters in the Bass curve equation for each of the products. The Bass equation is as follows:

(Eq. 12)
$$F(t) = \frac{1 - e^{-(p+q)t}}{1 + \left(\frac{q}{p}\right) e^{-(p+q)t}}$$

- *F(t) = the cumulative fraction of adopters*
- *p = coefficient of innovation*
- *q = coefficient of imitation*
- *t = elapsed time and tmax = when maximum adoption will occur (not in equation)*
- *Maximum adoption rate (not in equation)*

At least three of the above five metrics are needed in order to estimate the Bass curve, so the evaluators in California invented a surveying method whereby experts could log online and indicate three of those five variables in an online tool, which is shown below in Figure 20.

Figure 20. Online expert survey tool used to determine NOMAD in California appliance standards study

Source: Quantec 2007

3.2.1.3. Ex-post Evaluation Options

Ex-ante evaluations are quite common for appliance standards, since they help prove the case for energy efficiency, in both reducing a country or state's energy demand and likely saving consumers money in the process. These evaluations take on many assumptions about appliance performance in the field, however, and there is a growing interest in correcting *ex-ante* evaluations with field data to achieve more accurate savings estimates.

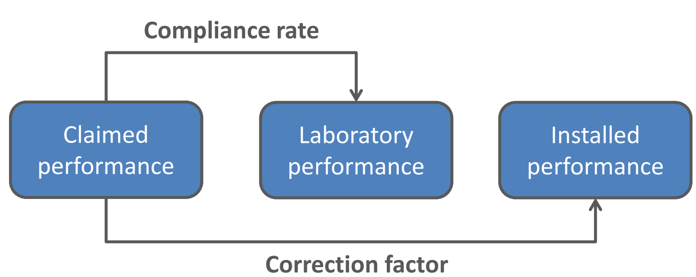


Figure 21: Relations of compliance rates and correction factors for performance data

Appliance efficiency may vary widely in the field due to a number of factors. For example, the UEC_e that is used in *ex-ante* evaluation is often based on engineering specifications as claimed or declared by the manufacturer. In many countries, these specifications are verified by third-party laboratory testing, but not all countries have product testing, and non-compliant products may still exist even where testing is done. Verification testing can be done on a sample of products to get a representative compliance rate, and test for how laboratory performance might differ from claimed performance as shown in Figure 21. Alternatively, metering measurements could be taken on an appliance once it is installed in a home or commercial building, and then installed performance could be compared with claimed performance.

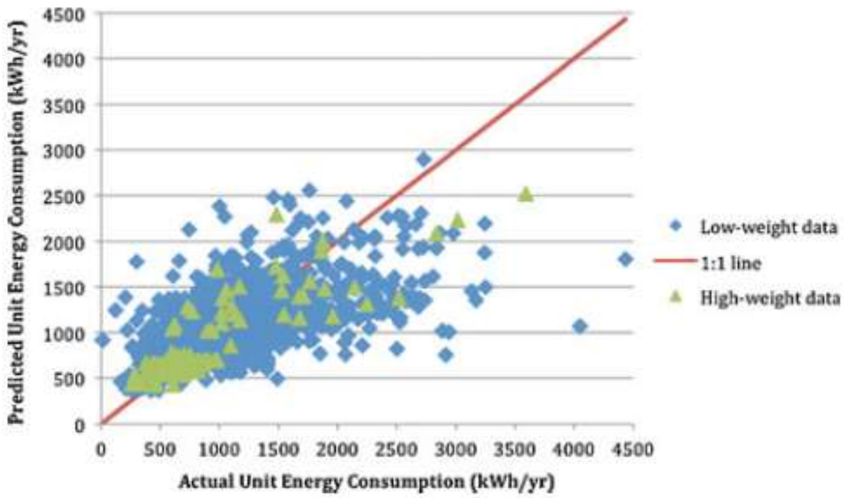


Figure 22. Predicted vs. actual UEC for refrigerators

Source: Greenblatt et al. 2011

In the case of refrigerators, test procedures are comprised of a static set of conditions, designed to, on average, produce similar energy consumption to that observed in the field. They are performed at 32°C or around 90°F – a high temperature compared to most operating conditions – and the refrigerator remains empty with no door opening and closing to simulate actual consumer activities. The artificially high ambient temperature compensates for the lack of door openings and (warm) food loading. The end result is that some actual (installed) UECs in individual households are higher or lower than predicted or claimed as shown in Figure 22. Researchers at LBNL have conducted studies on correction factors, also known as usage adjustment factors (UAF), with a range from 0.87 to 1.11, and an average of 0.99 to 1.01 (see Table 14 below). These types of studies are typically not conducted in *ex-post* evaluations as part of federal standards development (Greenblatt et al 2011).

In the case of the development of refrigerator standards, DOE developed UAF's using the DOE's Residential Energy Consumption Survey (RECS) data first to estimate the field energy usage of standard sized refrigerator/freezers and freezers on a representative sample of housing units. There was some concern about using RECS data to make these estimates, so instead field-metered electricity use data were collected from seven different studies with over 1,900 data points in all. The UAF was then calculated as the ratio of the field-metered annualized electricity use to laboratory test electricity use as below.

(Eq. 13) $$UAF_i = FEC_i / TEC_i$$

- UAF_i = usage adjustment factor specific to the field-metered data point;
- FEC_i = field-metered annualized electricity use;
- TEC_i = test energy consumption annual electricity use.

DOE performed regressions on a number of variables of potential interest in order to construct a function that predicts the UAF based on household and climate variables. Using the UAF specific to each household, product, and climate variable, they could then plug it in to a function to get the field energy consumption for each representative RECS household. This methodology is shown in Figure 23.

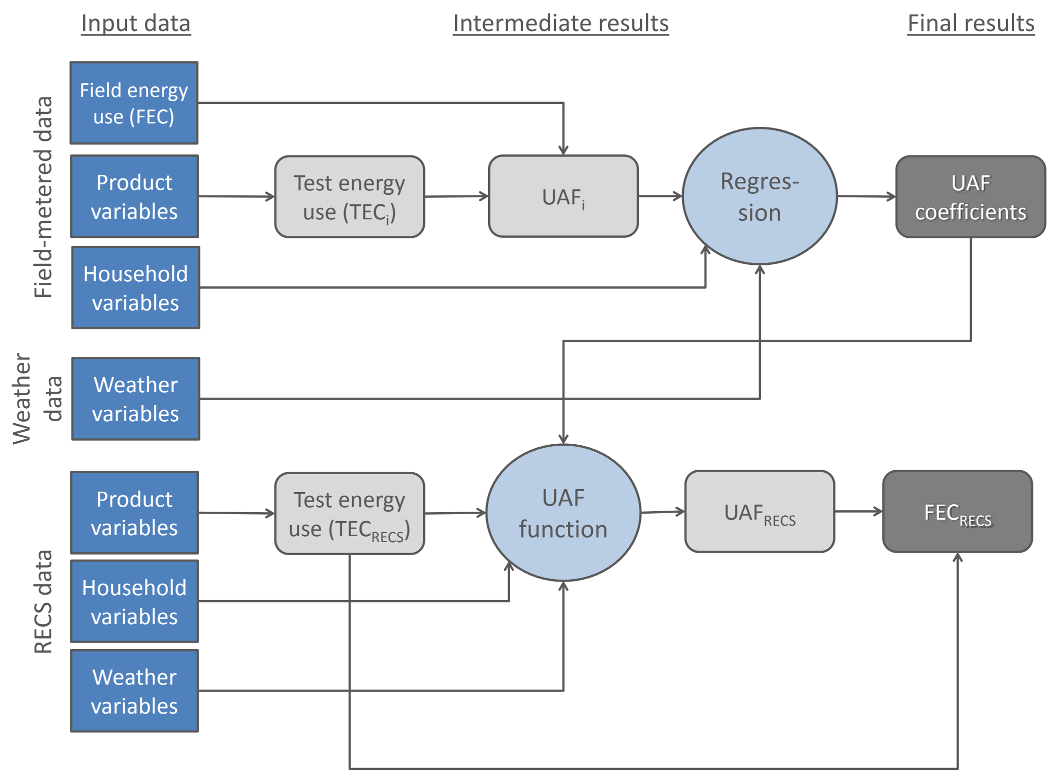


Figure 23. Flowchart for determining field energy consumption of different refrigerator product types

Source: DOE 2011a

For Australia's evaluation of S&L for refrigeration appliances, the evaluators concluded, based on limited test data, that the correction factor was about 0.9 for refrigerators and 0.75 for freezers, since the testing temperature was 32°C but actual operating conditions were 16-24°C. The energy use at 16°C can be 0.35-0.55 times that of the energy use at 32°C, so it is worthwhile to calculate these correction factors if budget allows for it. LBNL researchers have recommended that metering of appliances be incorporated into future versions of RECS, the data from which are used in many standards development and *ex-ante* evaluation calculations.

Like correction factors, adjustments for compliance rates have not been regularly used in evaluation studies to date. The California state standards evaluation study in 2007 did incorporate estimates of compliance (Quantec 2007). The evaluation did not look at verification testing of products in order to estimate compliance rates, but rather evaluated products against whether or not they had been registered with the CEC as part of the MEPS requirements. Manufacturers are obligated to submit an application with required information in order to sell their product legally in California. The evaluators took this list of compliant products and compared it to retail sales data (volumes and products) in order to settle an overall compliance rate. If a product was non-compliant (i.e., not properly registered with CEC) and it sold a high volume, this would have a significant impact on the appliance rate. In the end, the non-compliance rates of a number of products ended up being high (see Table 10), which will decrease the energy savings in the official standards evaluation. Based on the amount of data gathered, a certainty level was also assigned by the evaluators to the estimate (Quantec 2007).

Table 10. Summary of appliance non-compliance estimates in California state appliance standard evaluation

Appliance category	Estimated noncompliance rate	Certainty level of estimate
Televisions	41%	Medium
DVD players	57%	Medium
Residential pool pumps	15%	Medium
General service incandescents	27%	Medium
Metal halide luminaires	37%	Low
Walk-in refrigerators/freezers	0%	Medium
Pre-rinse spray valves	4%	High
Unit heaters and duct furnaces	44%	Low
Refrigerated canned/bottled beverage vending machines	63%	Low

Source: Quantec 2007

3.2.2. 3.2.2 Data requirements and sources

Data requirements for standards evaluation will depend on the scope and budget of the evaluation. For the Australian *ex-post* evaluation of their refrigerator S&L program, a fairly complete market dataset going back a number of decades established a strong foundation for the evaluation. Specifically, the Australians used a sales marketing company called GfK to gather the following types of data:

- **Stock of refrigerators by type by year by state:** Ownership of refrigerators is fairly uniform at a state level, however, there are significant variations in the ownership (stock) of separate freezers by state.
- **Sales of new refrigerators by group:** This is not often used directly in a stock model but these data are necessary to accurately weight the attributes of new products entering the stock by year, and information of total market sales provides an indirect indication of average product lifespan.
- **Sales by group to include attributes of volume/size, energy** and any features to allow an assessment of energy efficiency at a model level.
- **Distribution of refrigerators by climate zone** (although all evidence in Australia suggests that this effect is generally negligible – sales-weighted averages for all states are close to identical).
- **Correction functions** to convert 'test' data as declared by the manufacturer to actual consumption in the home, relating to climate zone (EE Strategies 2010)

Market sales for the most popular models were collected from 1993-2000, and market sales for *all* models were collected from 2001-2009 by GfK. Ownership and market saturation data were collected in 1994, 1999, 2002, 2005, and 2008 by the Australian Bureau of Statistics. Finally, a product registration database was used to associate sales data with other information on each model, including energy consumption, volume, defrost type, etc.

When this type of market sales data is not available, sales and stock data can be estimated using other parameters like urbanization, household income, and electrification as in the BUENAS model. Ownerships and market saturation is typically estimated using "snapshot" surveys which can be administered periodically. Good examples are the California Residential Efficiency Saturation Tool (CREST) and RECS. For CREST, the information in Table 11 is gathered roughly every five years on a sample of households that is large enough to accurately represent the California population.

Table 11. Data collected for California Residential Efficiency Saturation Tool

Energy uses and characteristics	Output	Cross tabulations
• Lighting • Heating and cooling systems • Refrigerators/freezers • Hot water heaters • Dishwashers • Clothes washers/dryers • Insulation • Windows • Consumer electronics • Etc.	• System type • Efficiency (UEC, EF, or similar) • Capacity • Age • Manufacturer date • Size • Etc.	• Utility • Type of residence (single-family, multi-family) • Age of residence • Climate zone • Residence size • Rent vs. own • Language • Other demographics…

Source: KEMA Inc. 2005

Table 12. Required and optional data requirements and sources for *ex-post/ex-ante* analysis of standards

Data type	Used in *ex-ante* or *ex-post*	Required or optional	Data source
Annual energy use per unit (UEC)	*Ex-ante, ex-post*	Required	Manufacturer test data
Existing stock	*Ex-ante, ex-post*	Required	Market data, government statistics
Market saturation (ownership, market shares)	*Ex-ante, ex-post*	Required	Market surveys
Lifetime or retirement function	*Ex-ante, ex-post*	Required	Manufacturer test data
Future shipment forecasts	*Ex-ante*	Required	Historic market data, government forecasts
Usage adjustment factor (UAF)	*Ex-ante, ex-post*	Optional	Metered test data
Naturally occurring market adoption (NOMAD)	*Ex-ante, ex-post*	Optional	Historic market data
Compliance rate	*Ex-post*	Optional	Metered test data
Real shipments/sales	*Ex-post*	Optional	Market data
Site-to-source energy conversion factors	*Ex-ante, ex-post*	Optional	Power plant energy data
Emission factors	*Ex-ante, ex-post*	Optional	Power plant emission data

Table 12 outlines the most commonly used data types that were described in Section 3.2.1. About half of the data types are required to perform an *ex-ante* or *ex-post* evaluation, while many of the optional data types are *ex-post* adjustments made to *ex-ante* estimates, such as UAF's, compliance rates, and real shipments/sales data. Market data and surveys, as well as manufacturer test data, account for the main data sources for all required data types. Metered test data and historic market data can help improve evaluation estimates.

3.2.3. Example Calculations

3.2.3.1. Ex-ante Evaluation

This section demonstrates some example calculations for *ex-ante* and *ex-post* evaluations based on selected studies. For *ex-ante* evaluation, the main example will be from the ex-ante evaluation done as part of the national impact analysis conducted by DOE in setting new refrigerator standards in the U.S. The first step in performing the *ex-ante* evaluation is setting the stock of products according to product type and efficiency, and tracking how that stock changes over time. The failure rate equation described in Section 3.2.1 is used to dictate replacement units. Figure 24 describes a typical distribution for the survival of an appliance. The average life of a refrigerator is a little over 15 years. The failure rate will increase every year beyond a given threshold (5-6 years for refrigerators). By 30 years of age, about 95% of the refrigerators will have retired.

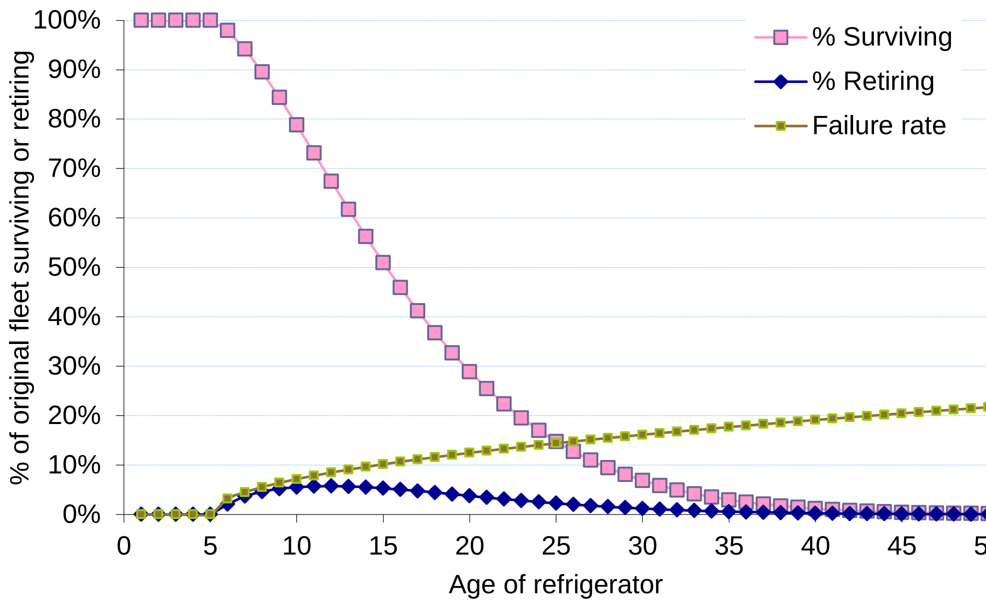


Figure 24. Failure, retirement, and survival rates for refrigerators as a function of age in years

Source: DOE 2011b

This turnover rate is critical for determining when consumers will purchase a new appliance that is subject to the new standard at a higher efficiency level. These purchases will be documented as replacement units in the stock model. The model will also incorporate estimates for new construction (based on government housing statistics and projections) and second refrigerator purchases (out of scope for this study). New shipments will be the sum of replacements, new construction, and conversions to second refrigerators. Table 13 shows an example of these datasets as modeled by DOE for their national impact analysis *ex-ante* evaluation of proposed standards. Figure 25 shows the end result in graph form, as compared with historical shipment data.

Table 13: New shipments as broken down by replacement units, new construction, and second refrigerators

Year	New Shipment	Total Stock	Market Segments			
			Replace-ment Units	Shipments to Housing Starts	Conversion to 2 refrig	Modeled Total Shipment
2005	11.134	140.439	6.594	2.807	*1.26*	10.659
2006	11.078	144.745	6.772	2.476	*1.28*	10.523
2007	10.402	148.192	6.955	1.894	*1.26*	10.113
2008	9.314	150.357	7.149	1.294	*1.19*	9.630
2009	9.223	152.227	7.353	0.801	*1.07*	9.223
2010	9.865	154.512	7.580	1.218	*1.07*	9.865
2011	10.928	157.618	7.822	1.957	*1.15*	10.928
2012	11.619	161.167	8.070	2.272	*1.28*	11.619
2013	12.017	164.881	8.303	2.356	*1.36*	12.017
2014	12.283	168.669	8.496	2.380	*1.41*	12.283
2015	12.567	172.582	8.654	2.475	*1.44*	12.567
2016	12.842	176.625	8.799	2.570	*1.47*	12.842
2017	13.063	180.732	8.956	2.602	*1.51*	13.063
2018	13.298	184.900	9.130	2.637	*1.53*	13.298
2019	13.559	189.137	9.322	2.679	*1.56*	13.559
2020	13.795	193.406	9.526	2.680	*1.59*	13.795

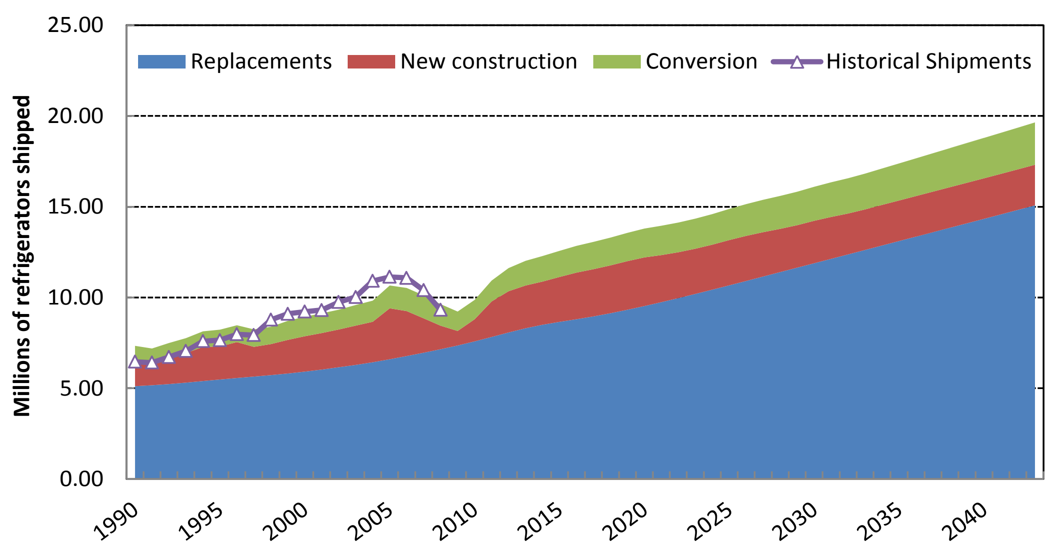


Figure 25. Modeled refrigerator shipments by replacements, new construction, and conversion as compared with historical shipment data

Source: DOE 2011b

Once the stock has been set, the next step is to calculate the UEC for baseline and proposed standards levels within given product classes. According to the method described in Table 8 in Section 3.3, projected efficiency distributions are multiplied by proposed UEC_e levels to get the shipment weighted UEC. In this case, the highest level of standard is being tested, so this shows the standard taking effect in 2014, and the UEC_e changing from 519 kWh/year to 347 kWh/year. The 2013 weighted average energy use is calculated by multiplying the market shares on the left by their associated UEC in the right table and summing the parts to arrive at the 519 kWh/year seen below. Since the market share for this standards case looks at a complete shift to the highest efficiency product in 2014 when the standard is implemented, the UEC drops to 347 kWh/year.

RRF with manual defrost							PC1
Level	**0**	**1**	**2**	**3**	**4**	**5**	**6**
2005	81%	6%	13%	0%	0%	0%	0%
2006	81%	6%	13%	0%	0%	0%	0%
2007	81%	6%	13%	0%	0%	0%	0%
2008	76%	2%	19%	3%	0%	0%	0%
2009	76%	2%	19%	3%	0%	0%	0%
2010	76%	2%	19%	3%	0%	0%	0%
2011	76%	2%	19%	3%	0%	0%	0%
2012	76%	2%	19%	3%	0%	0%	0%
2013	76%	2%	19%	3%	0%	0%	0%
2014	0%	0%	0%	0%	0%	0%	100%
2015	0%	0%	0%	0%	0%	0%	100%
2016	0%	0%	0%	0%	0%	0%	100%

		PC 1	
Level	**Efficiency Level**	**Efficiency**	**UEC**
		% over base	kWh/year
0	Baseline Efficiency	0%	539
1	Level 1	10%	485
2	Level 2	15%	458
3	Level 3	20%	431
4	Level 4	25%	404
5	Level 5	30%	377
6	Level 6	36%	347

Base Case Weighted Average Energy Use

	Annual Energy Use						
PC	**1**	**2**	**3**	**4**	**5**	**6**	**7**
	kWh/yr	*kWh/yr*	*kWh/yr*	*kWh/yr*	*kWh/yr*	*kWh/yr*	*kWh/yr*
2005	524	524	524	741	572	524	741
2006	524	524	524	741	572	524	741
2007	524	524	524	741	572	524	741
2008	519	519	519	704	564	519	704
2009	519	519	519	704	564	519	704
2010	519	519	519	704	564	519	704
2011	519	519	519	704	564	519	704
2012	519	519	519	704	564	519	704
2013	519	519	519	704	564	519	704
2014	520	520	520	716	556	520	716
2015	520	520	520	716	555	520	716
2016	520	520	520	716	554	520	716

Figure 26. Shipment weighted UEC calculation for national impact analysis; Note: Product category (PC) 1 is shown in red. PC's #2-7 are calculated using different UECs and sales weightings

Source: DOE 2011b

Next, the UEC is corrected with a UAF to get a more accurate estimate based on how refrigerators operate in the field versus in the test laboratory. Table 14 shows the mean UAF from the final rule analysis calculated using field-metered data. DOE also found that there was a change in UAF over time as shown in Figure 27. Newer products perform much more efficiently than in testing (UAF<1), while older products (20 years old or more) operate less efficiently than the test procedure (UAF>1).

Table 14. Average and range of usage adjustment factors for different refrigerator product classes

Product class	Sample size	Mean UAF – average (range)
Top-mount refrigerator-freezer	2,303	1.00 (0.88 to 1.11)
Bottom-mount refrigerator-freezer	2,303	0.99 (0.87 to 1.10)
Side-by-side refrigerator freezer	1,026	1.01 (0.90 to 1.11)

Source: DOE 2011a; Note: range for final rule analysis is for year 1 to year 20 of product lifetime.

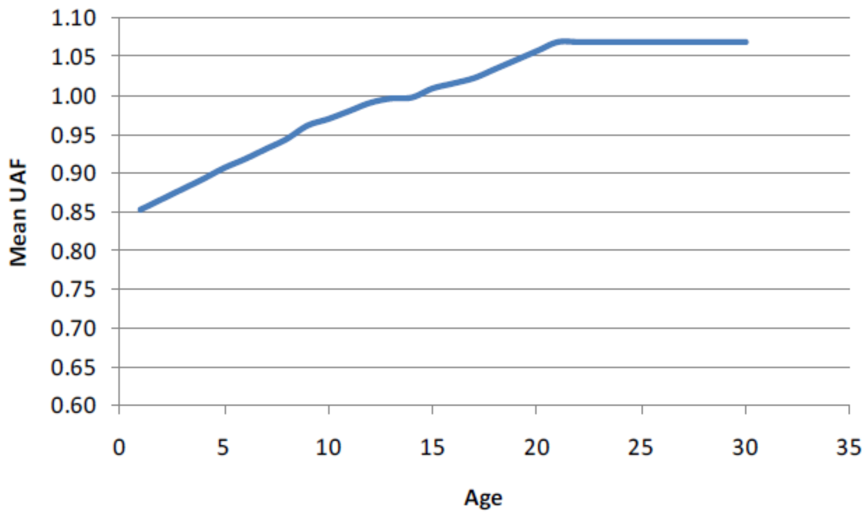


Figure 27. Mean usage adjustment factor as a function of age of product

Source: DOE 2011a

The final step is to calculate the gross annual energy savings according to the following equations:

(Eq. 14 and 15) $$\mathrm{NES_y} = \mathrm{AEC_{BASE}} - \mathrm{AEC_{STD}} \quad | \quad AEC_y = \sum STOCK_V * UEC_V$$

which can be simplified to:

(Eq. 16) $$NES_y = \sum STOCK_V * \Delta UEC_V$$

where ΔUEC_V is the difference between the standards and the base case shown in Figure 28. According to the survival function and the new shipments function, the surviving stock of a specific vintage year V will be calculated and multiplied by the change in UEC between the standard level and the base case, which is the UEC for the product that was purchased in year V. The equation will also use the UAF as described previously.

UEC (kWh/yr)	Base case	Standards case
2005	524	524
2006	524	524
2007	524	524
2008	519	519
2009	519	519
2010	519	519
2011	519	519
2012	519	519
2013	519	519
2014	520	347
2015	520	347
2016	520	347

Figure 28. UEC for base case and standards case, visible starting in 2014

Source: DOE 2011b

3.2.3.2. Ex-post Evaluation

Once gross savings have been established, there is an option to perform further optional calculations in order to get a more realistic estimate. The following examples are drawn from a California state-level study on new standards that California implemented for appliances and electronics not covered by federal standards; in this case, a standard for standby power usage for televisions. The example begins with gross energy savings and then incorporates estimates on NOMAD, compliance rate, and normally occurring standards development. Table 15 lists the main parameters needed for this *ex-post* evaluation.

Table 15. Parameters for *ex-post* evaluation study on California television standard

Parameters		
Gross energy parameters	Gross energy savings	68 GWh
	True up for actual installations	1
Bass curve parameters (NOMAD)	Max penetration	83%
	Starting year	2000
	p	0.07
	q	0.17
Compliance rate parameters	Assumed non-compliance rate	41%
Other parameters	Assumed code update rate	6 years

Source: Quantec 2007

Efficiency standards are assumed to be updated periodically according to Title 20, California's appliance standards act. In the case of TV's, six years is the assumed code update rate.

To determine the non-compliance rate, the evaluators visited 13 stores and took inventory to find if the product was or was not in the CEC database; it was listed as non-compliant if it was not in the database. The evaluators were able to identify a total number of models sold in store, but could only get inventory data for a subset of those models. A total inventory was taken on nearly 3,000 televisions for which there were inventory data, and it was found that noncompliance was at 41%, as shown in Table 16.

Table 16. Noncompliance rate for televisions

Data used	Total Count	Met standard	Noncompliant	Noncompliance %
Total models	876	402	474	54%
Only models with inventory data	537	236	301	56%
Inventory	2,943	1,174	1,199	**41%**

Source: Quantec 2007

Lastly, the parameters for the Bass curve and NOMAD rate need to be estimated. Experts responded via the online tool described in Figure 20. The responses were tallied and an average was calculated, as shown in Figure 29. From this average, the innovation and imitation coefficients can be determined, and the NOMAD curve is calculated.

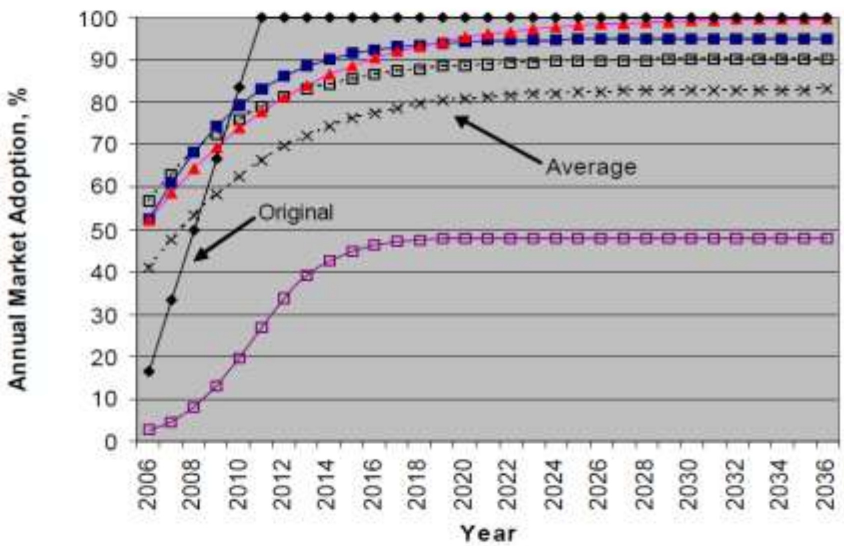
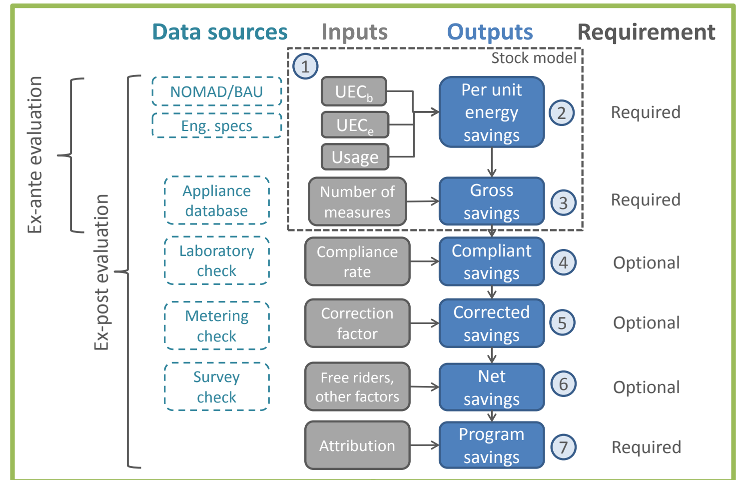


Figure 29. Compliant television market adoption estimates

Figure 30 on the following page further details the calculations and steps taken in this *ex-post* evaluation calculation. As seen in the spreadsheet, the annual net energy savings is drastically different than the annual gross energy savings. Both the NOMAD and normally occurring standards have a huge impact on the standard. Specifically, the methodology assumes that once the standard is updated, there will be no new net savings from that standard. Instead, savings will be associated with the new, updated standard. The NOMAD and non-compliance adjustments are also significant in reducing the gross savings amount. It should be noted, however, that the methods for both of those adjustments were quite crude for this study. The non-compliance adjustment was based on product inventories and certifications as opposed to actual test laboratory (via product verification) or metered test data, while the estimation of NOMAD was based on surveys of industry experts.

	Steps	1. Annual gross energy savings (GWh)	2. True-ups for actual installations	3. Market penetration model	4. Naturally occurring market adoption	5. Non-compliance adjustment	6. Normally occuring standards adoption	7. Total adjustments	8. Annual net energy savings (GWh)
1	2006	68	68	0.41	28	16	0	44	24
2	2007	68	68	0.47	32	15	0	47	21
3	2008	68	68	0.53	36	13	0	49	19
4	2009	68	68	0.58	40	12	0	51	17
5	2010	68	68	0.63	43	10	0	53	15
6	2011	68	68	0.66	45	9	13	68	0
7	2012	68	68	0.70	47	8	12	68	0
8	2013	68	68	0.72	49	8	11	68	0
9	2014	68	68	0.74	51	7	10	68	0
10	2015	68	68	0.76	52	7	10	68	0
11	2016	68	68	0.78	53	6	9	68	0
12	2017	68	68	0.79	53	6	9	68	0
13	2018	68	68	0.80	54	6	8	68	0
14	2019	68	68	0.80	55	5	8	68	0
15	2020	68	68	0.81	55	5	8	68	0



Step number	Step number	Methodology description
1	3	Use annual gross energy savings estimate
2		True up to see if the measure was actually installed. This is more commonly used in the case of lightbulb giveaway programs.
3	1	Use bass curve equation and bass curve parameters to estimate naturally occuring market adoption.
4	1	Multiply gross savings by naturally occuring market adoption
5	4	Multiply gross savings by non-compliance rate
6	1	Assume that once the code is updated (according to code update rate), there will be no savings.
7	6	Add up savings lost due to naturally occuring market adoption, non compliance, and normally occurring standards.
8	6	Subtract total adjustments from gross savings to arrive at net savings.

Figure 30. Spreadsheet, framework (see also Figure 14), and calculation steps for *ex-post* evaluation of standards from gross to net savings estimation

Source: Adapted from Quantec 2007

3.3. Impact Evaluation of Appliance Labeling Programs

Impact evaluation methodologies for labeling are very similar to standards evaluation methodologies. The main difference is that instead of dividing sales up into appliances that meet the standard and appliances that do not meet the standard, a more precise division is needed. For instance, in categorical labeling, often three or five labeling categories are used so sales will need to be divided up into as many categories plus an additional category for any appliances that do not meet even the basic standard. For voluntary certification labeling such as ENERGY STAR, three categories are needed – appliances that do not meet the minimum standard, appliances that meet the minimum standard, and appliances that meet the higher categorical labeling standard. Similarly, other evaluation methodologies such as compliance adjustments or baseline setting exist in labeling evaluation as they do in standards evaluation. So the following section will focus mostly on the establishment of efficiency scenarios according to a labeling program and the efficiency classes that program creates.

3.3.1. Review of Existing Evaluation Methodologies

Two tools and accompanying case studies show how efficiency scenarios are formulated and implemented in evaluating the net energy savings impact of categorical information labels. In addition, a U.S. case study of the model is used to evaluate the energy and emissions reduction impacts of the ENERGY STAR certification labeling program.

3.3.1.1. EU Estimation Tool for the National Effects of EU Ecodesign Standard and Energy Labeling Programs

In 2012, researchers at the Danish and Swedish Energy Agencies developed and introduced an estimation tool to quantify the national, rather than EU region-wide, effects of the EU Ecodesign standards[3] and EU energy information labeling program (Larsen et al. 2012). This tool uses a standard bottom-up, Excel-based stock turnover model to account for the energy consumption and potential energy savings given different conditions (e.g., climate, usage) in the Nordic countries. This tool serves as a simplified alternative to more complex tools (such as BUENAS) and has been used to estimate the Ecodesign standards and E.U. energy labeling effects for televisions and lighting in Denmark and Sweden.

This simplified E.U. labeling effects estimation tool for televisions requires inputs for the following key parameters:

1). <u>Equipment sales by energy efficiency class and by product subgroups</u>: television sales data were collected from questionnaires by European market research companies on historical and recent sales distribution by E.U. Energy Label efficiency class (A+++ through G). The sales data were further divided into five common product subgroups by differences in technology (e.g., LCD, LED, Plasma) and television screen size. Figure 31 shows a matrix of the sales distribution over time by efficiency class for one of the five product subgroups (LCD-40[4]).

3 The EU Ecodesign program sets mandatory requirements based on lifetime performance criteria with emphasis on energy consumption and environmental aspects of the non-use phases of energy-using products. This program has a broader scope than minimum energy performance standards.

4 LCD-40 refers to 40-inch liquid crystal display televisions.

	1995	1996	1997	1998	1999	2000	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010
G	65	60	55	50	45	40	35	30	25	20	17	27	22	28	19	12
F	35	35	35	35	35	35	32	29	25	22	24	43	40	33	23	21
E	0	5	10	15	20	25	30	35	40	42	40	17	24	20	27	18
D	0	0	0	0	0	0	3	6	10	15	17	11	14	19	24	27
C	0	0	0	0	0	0	0	0	0	1	2	2	0	1	7	21
B	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1
A	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
A+	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
A++	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
A+++	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0

Figure 31. E.U. Estimation Tool Sales Distribution (percent) by Efficiency Class for Product Subgroup LC-40

Source: Larsen et al. 2012

2). Lifetime: the lifespan of equipment, defined as the age at time of replacement or disposal, influences the stock accounting model and the retirement function. For televisions, a normal distribution of retirement was used with an assumed lifetime of 7 years and standard deviation of 2 years.

3). Baseline energy consumption of product subgroups: a baseline UEC was determined for each of the product subgroups by applying the formula for calculating energy consumption in the Ecodesign Directive to an assumed average size and assumed daily hours of usage for each subgroup. The assumed average size was derived from questionnaire data from market research companies and forecasting tools, while the assumed hours of usage was taken from national data.

Figure 32 shows the key technical inputs needed to calculate the baseline per UEC for a given product subgroup (LCD-40).

Longevity	7	2	Years	**ON**	1750	Hours/year
				Standby	2071	Hours/year
Formula	E=Ton*(20+1,12*4,3224*Size)+Pstdb*Tstdb					
EEI reference		LCD-40	29	Std power		Watt
		Size	23	Standby	1	Watt
		Std cons.	233.5	kWh/year		
			132.2378			
Eco-design criteria - defaults						
	kWh	EEI				
2011	233.5	1	New 2012 crit.:			
2012	170.4	0.729665	170.4			
2014	170.4	0.729665				
2028	170.4					
2050						

Figure 32. E.U. Estimation Tool's Technical Inputs to Calculating Base Energy Consumption

Source: Larsen et al. 2012

4). Projected sales: future television sales were derived using a simple forecast of total sales assuming a linear trend combined with expert knowledge and an assumed natural development of 2% increase in efficiency per year in sales distribution. The basis for the projected sales is expert knowledge, combined with pre-determined linear trends driven by Gross Domestic Product (GDP) growth and other third-party sales forecasts.

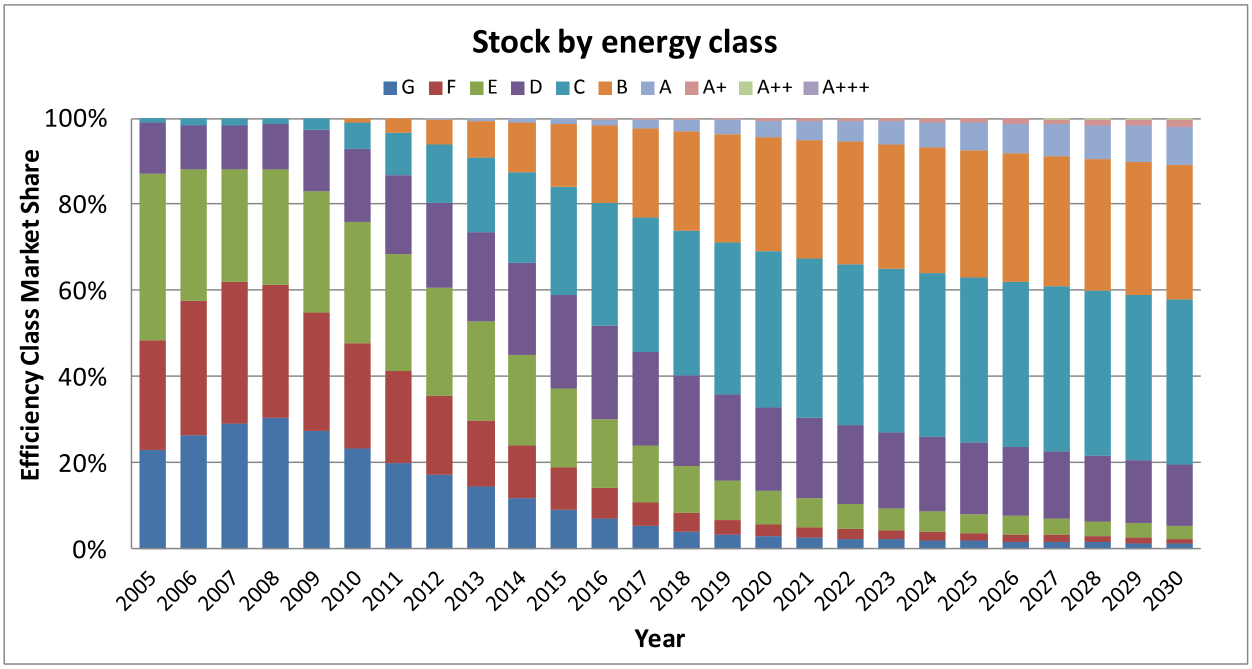


Figure 33. E.U. Estimation Tool Project Sales by Energy Class

Source: Larsen et al. 2012

Given the above data inputs and stock model parameters, two key calculations were performed in the model. First, the total television stock per energy class and product subgroup is calculated as the sum of all vintage of television sales given the lifetime and retirement function. Second, the base energy consumption of the entire stock for a given energy class and product subgroup in a given year is calculated as the UEC multiplied by that year's total stock.

These two key calculations are applied to two scenarios to differentiate between the total energy consumption under MEPS and under the E.U. Energy Labeling program.

The MEPS scenario limits television sales to the allowed efficiency classes under current legislation and automatically shifts inefficient classes that are eliminated by the legislation to the next higher efficiency class (E.g., shifting from class G to F). This scenario also accounts for natural efficiency gains outside of MEPS by incorporating the assumed natural 2% increase in efficiency per year in sales distribution.

The labeling scenario assumes a higher sales shift of 25% per year towards higher efficiency classes as a result of the labeling program impact, based on data collected for white goods in Denmark in the 1990s.

The labeling scenario is built in parallel to MEPS, so that sales shifts already simulated by the MEPS scenario will not be affected by labeling.

The results for the energy impacts of MEPS and the combined effects of MEPS and labeling can be calculated by summing up the energy consumption results for all five subgroups in each of the two scenarios.

3.3.1.2. LBNL LEAP Model for Estimating China Energy Label Effect on Refrigerators

In 2008, LBNL developed a bottom-up stock turnover model using the Long-range Energy Alternatives Planning (LEAP) end-use energy accounting modeling platform from the Stockholm Environment Institute to estimate the energy and emissions reduction impact of the China Energy label on refrigerators. The overarching methodology used in this study is similar to the EU estimation tool, with the key difference being that the stock turnover model was implemented using LEAP software rather than Excel spreadsheets.

The LEAP model for refrigerators was developed using the following parameters and data inputs for refrigerators:

1. Historical and recent sales split by label energy efficiency grades and product subgroups: recent sales data by model were provided by LBNL collaborators at the China National Institute of Standardization in China. While the China Energy Label for refrigerators have 5 efficiency grades (Grades 1 through 5), some efficiency grades were collapsed and simplified into three efficiency classes of Ordinary (Grade 5), Efficient (Grades 3 and 4) and Highly efficient (Grades 1 and 2) in the model.
2. Lifetime: the age at the time of replacement or disposal was assumed to be 12 years for all units, rather than using an assumed average lifetime. Due to limited data on actual retirement trends, the retirement function was implemented as a simple step function where all equipment will be retired at the end of its 12th year of life, rather than a normal distribution retirement function with some units retiring before and after the 12th year.
3. Product subgroups by size: three common refrigerator sizes were implemented in LEAP to reflect consumer preferences in the refrigerator market. Historical and project shares of refrigerators by each of the three sizes are incorporated into the model.
4. Projected sales: future total sales of refrigerators were projected assuming a linear growth trend and calibrated to historical sales.
5. Marginal energy intensity: the average marginal energy intensity or UEC for refrigerators is calculated based on efficiency distribution and size categories. Table 17 shows the range in assumed UEC values for refrigerators by size and efficiency class.

Table 17. LBNL LEAP Model Assumed Refrigerator UEC Values

kWh/year	2008			2012		
	170-liter	220-liter	270-liter	170-liter	220-liter	270-liter
Ordinary	351	391	436	307	342	382
Efficient	281	313	349	246	274	306
Highly efficient	228	254	283	200	222	248

Source: Fridley et al. 2008

With the above parameters and inputs, the LEAP model then performs two key calculations to determine the total stock and total electricity consumption for a given efficiency grade and product subgroup. The total stock for a given year is derived by LEAP using the sum of vintage of total sales and assumed lifetime. Total electricity consumption for a given stock of refrigerators is calculated using the following formula:

$$Electricity\ (TWh) = \ Marginal\ Energy\ Intensity\ \left(\frac{\frac{kWh}{year}}{unit}\right) \times Refrigerator\ Stock\ (units)$$

To calculate the total electricity consumption and potential savings from the China Energy Label for refrigerators, the following three scenarios were implemented in LEAP.

1. Baseline Scenario: the baseline scenario assumes a frozen market distribution of refrigerators and frozen efficiency at the base year level with no future efficiency improvements. This baseline is intended to serve as the counterfactual scenario for measuring the impact of MEPS and energy labeling.
2. MEPS Scenario: this scenario is based on the baseline scenario but differs in assuming that there is a shifting of the efficiency grades of label by 10% in 2012 as tightened MEPS become effective. This leads to a rise in the ordinary efficiency class of refrigerators and a decline in super-efficient class of refrigerators due to tightened MEPS requirements and the absence of further market shifts. Market distribution by efficiency class is then assumed to remain frozen from 2012 to 2020.
3. Label Scenario: this scenario is based on the MEPS scenario but differs in assuming an additional market distribution shift effect that results from the impact of the Energy Label for refrigerators. In 2012, small market shift effects from the labeling impact are modeled in addition to the 10% market shifting effect from MEPS. The market shift effect as a result of the energy impact is assumed to be analogous to the dynamic market shift experiences of the EU label during the initial years of 1992 to 1996. From 2012 to 2020, there are continued market distribution shifts towards the more efficient classes as a result of the labeling impact. This is again based on the EU labeling experience through 2003, where the majority of market share is assumed to reach super-efficient and efficient classes by 2020 in China.

The key differences between the MEPS and label scenario can be seen by comparing the market shares by efficiency classes over time shown in Table 18.

Table 18. LBNL Refrigerator Study Market Shares by Efficiency Class

MEPS Scenario	2008	2012	2020
Super-Efficient	27%	19%	19%
Efficient	61%	61%	61%
Ordinary	13%	20%	20%
Label Scenario	**2008**	**2012**	**2020**
Super-Efficient	27%	33%	79%
Efficient	61%	60%	20%
Ordinary	13%	7%	2%

Source: Fridley et al. 2008; Note: Shares do not always add to 100% owing to independent rounding

The formulas for calculating the electricity savings of MEPS and labeling in LEAP are then as follows:

MEPS Savings = Baseline Total Electricity Consumption – MEPS Scenario Total Electricity Consumption

Labeling Savings = MEPS Scenario Total Electricity Consumption – Label Scenario Total Electricity Consumption

Figure 34 and Figure 35 depict the results for total electricity consumption and electricity savings from labeling from the refrigerator LEAP model.

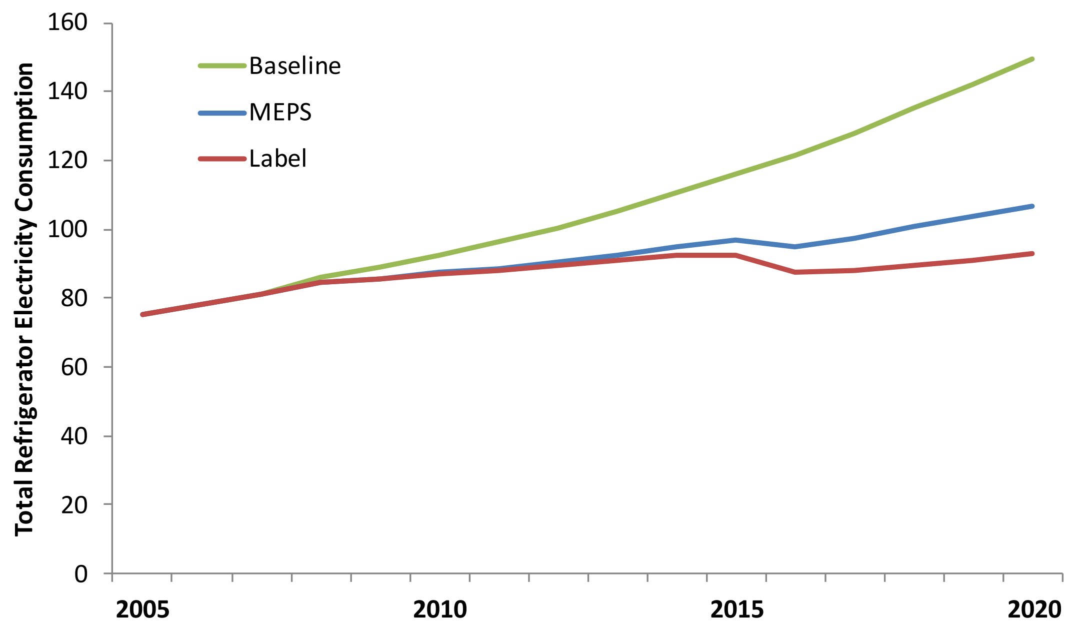


Figure 34. LBNL Refrigerator Study Total Electricity Consumption by Scenario

Source: Recreated based on data from Fridley et al. 2012

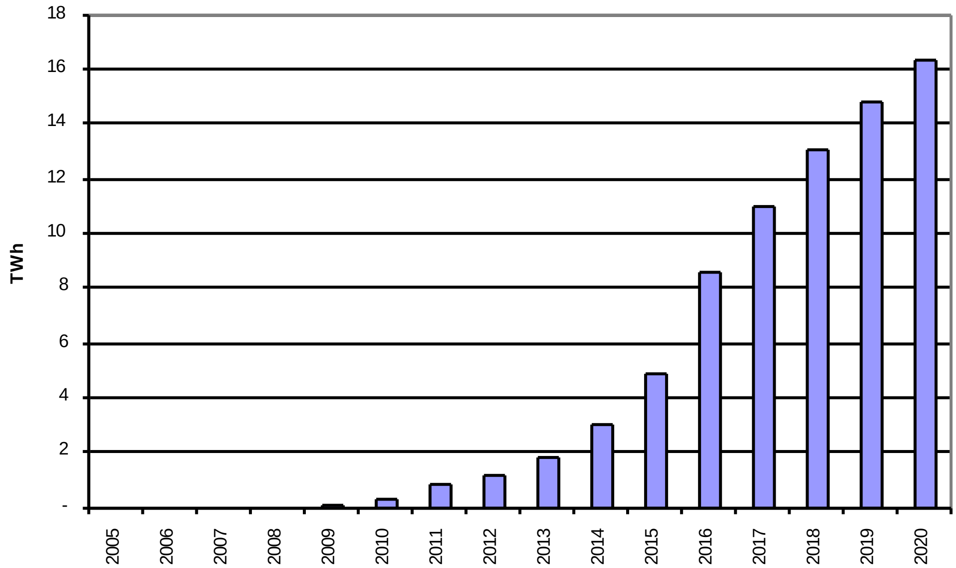


Figure 35. LBNL Refrigerator Study Labeling Electricity Savings

Source: Fridley et al. 2008

3.3.1.3 ENERGY STAR Program Savings Evaluation

As part of the U.S. ENERGY STAR voluntary energy efficiency endorsement labeling program, periodic evaluations of benefits from the program are conducted for all products covered by the labeling program. In 2010, an evaluation of current and projected benefits including energy and carbon savings from the 2009 ENERGY STAR program was conducted by LBNL on behalf of the U.S. EPA and U.S. DOE.

Similar to the two previous models for evaluating the impacts of energy information labels, the ENERGY STAR program evaluation of benefits is done using a bottom-up methodology. But unlike the previous two tools that conducted evaluations of the labeling impact for specific products, the ENERGY STAR evaluation uses product-specific inputs and impacts to characterize each ENERGY STAR product individually and then sums up the individual impacts of all products to derive the total program impact. The ENERGY STAR methodology is also based on specific assumptions and inputs for product sales, UEC, and annual energy savings.

First, data on total U.S. sales of products included in the ENERGY STAR program as well as market segments of sales related to ENERGY STAR products are collected. To differentiate between the effects of the labeling program on market efficiency improvements, sales data are collected and estimated for five different types of products, using a market segmentation analytical approach shown in Figure 36.

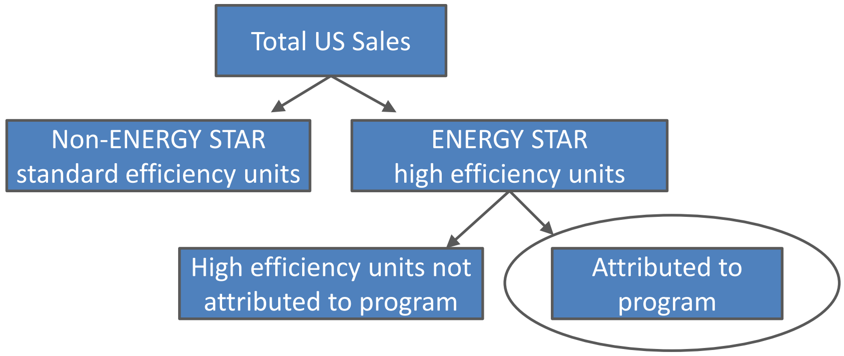


Figure 36. Market Segmentation of ENERGY STAR Products

Source: Homan et al. 2010

Total U.S. sales data for all units are taken from industry reports, while ENERGY STAR sales data for ENERGY STAR high efficiency units are obtained from ENERGY STAR manufacturers and partners, which are obligated to report sales data as part of the program. Sales of the Non-ENERGY STAR standard efficiency units are then calculated as the difference between total U.S. sales and ENERGY STAR product sales. At the next level, ENERGY STAR product sales are further divided into two subgroups to account for market effects that can and cannot be attributed to the program. The first sub-group of ENERGY STAR products sold is high efficiency units not attributable to the program, or what are known as BAU high efficiency units because these units would have been manufactured for purchase by a small proportion of consumers even if the label did not exist. This subgroup of products is determined through market share analysis of models that met the ENERGY STAR specifications before the specification went into effect, using data reported as part of energy consumption test results from partner manufacturers during the specification development process. The remaining subgroup of ENERGY STAR products sold (i.e., high efficiency units attributed to the program) is then the basis for evaluating the impact of the labeling program.

Another important component of the evaluation methodology for quantifying the ENERGY STAR program impacts is the determination of UEC for standard efficiency and high efficiency ENERGY STAR units. Two forecasts are used to evaluate the total energy consumption of products covered by the program: a BAU forecast which represents standard efficiency units and high efficiency units not attributable to the program and an ENERGY STAR forecast which represents high efficiency units attributed to the program. For the BAU forecast, individual UEC and market shares for both standard efficiency and high efficiency units are used to estimate an average UEC. For the ENERGY STAR forecast, the UEC is estimated using ENERGY STAR product specifications for the current year. The unit energy savings attributed to ENERGY STAR products is then calculated as the difference between the BAU UEC and the ENERGY STAR UEC.

Combined with stock accounting based on tracking of different vintages of products and a retirement function based on average product lifetime, the unit energy savings is used to calculate the annual energy savings of the ENERGY STAR program and associated financial and environmental benefits using the follow equations (Homan et al. 2010):

$$AES_t = \sum_{n=t-L}^{t} X_n UES_n$$

$$Energy\ Bill_t(undiscounted) = AES_t P_t$$

$$Carbon\ savings = AES_t C_t$$

- X_n = number of ENERGY STAR units sold in year *n* due to the program
- UES_n = unit energy savings of ENERGY STAR units sold in year *n* (in kWh or MBtu)
- L = product lifetime
- AES_t = aggregate annual energy savings in year *t* (in kWh or MBtu)
- P_t = energy price in year *t* (in $/kWh or $/MBTu)
- C_t = carbon emissions factor in year *t* (in kgC/kWh or kgC/MBtu)

Table 19 shows the results of the ENERGY STAR evaluation methodology for estimating achieved savings for 2009. Similar results are produced for estimating potential savings for future years using projected parameters such as energy prices and marginal electricity carbon emission factors.

Table 19. Achieved Annual Savings from ENERGY STAR Program in 2009

Program	Equipment Type	Primary Savings	Energy Bill Savings, Discounted	Carbon Emissions Avoided	Conservation Load Factor	Peak Load Savings
		Trillion Btu	Million $2008	MtC		GW
Office Equipment	Computers	40	390	0.7	0.0089	29
	Servers	0.2	1.9	0.0035	1	2.8
	Displays (Monitors)	310	2900	5.4	1.4	2.8
	Fax	2.3	23	0.04	1	0.018
	Copier	23	210	0.4	4.6	0.071
	Multifunction Device	38	350	0.66	1.1	0.36
	Scanner	1.2	12	0.021	0.76	0.011
	Printer	77	730	1.4	4	0.25
	Professional Displays	0	0	0	0.42	0
	Subtotal	490	4600	8.5	1.5	3.7
Consumer Electronics	Digital Picture Frames	0	0	0	1	0
	TVs	130	1300	2.3	1	1.4
	VCRs	4.2	43	0.074	1	0.044
	TV/VCR/DVD	8	82	0.14	1	0.084
	DVD Player	8.7	89	0.15	1	0.091
	Audio Equipment	9	92	0.16	1	0.094
	Telephony	22	220	0.38	1	0.23
	Digital TV Adapter	5.8	60	0.1	0.69	0.089
	Set-top Box	12	120	0.21	1	0.13
	External Power Supplies	68	660	1.2	1	0.72
	Battery Charging Systems	1.6	16	0.028	1	0.017
	Subtotal	270	2700	4.7	0.99	2.8
Heating & Cooling	Furnace (Gas or Oil)	49	550	0.75	-	-
	Central Air Conditioner	32	320	0.55	0.15	2.2
	Air-Source Heat Pump	30	310	0.52	0.15	0.78
	Geothermal Heat Pump	13	130	0.22	0.15	0.1
	Boiler (Gas or Oil)	4.4	64	0.074	-	-
	Programmable Thermostat	0	0	0	0.15	0
	Unitary HVAC	54	490	0.94	0.15	3.7
	Subtotal	180	1900	3.1	0.18	6.8
Residential	Fixtures	98	1000	1.7	1	1

and Commercial Lighting	CFLs	370	3800	6.5	1	3.8
	Exit Sign	4.1	38	0.072	1	0.043
	Decorative Light Strands	0.66	6.8	0.012	1	0.0068
	Traffic Signal	9.9	91	0.17	1	0.1
	Subtotal	490	5000	8.5	1	5
Residential Appliances	Room Air Conditioners	20	210	0.36	0.15	1.4
	Dehumidifiers	9.3	96	0.16	0.38	0.26
	Air Cleaners	4.6	47	0.081	1	0.048
	Exhaust Fans	1.9	20	0.034	1	0.02
	Ceiling Fans	1.5	16	0.026	1	0.016
	Dishwashers	39	410	0.65	0.77	0.38
	Refrigerators	27	280	0.47	0.95	0.3
	Clothes Washers	44	460	0.73	0.65	0.52
	Subtotal	150	1500	2.5	0.44	3
Commercial Appliances	Water Coolers	14	130	0.24	0.7	0.22
	Commercial Refrigeration	8.9	82	0.16	0.95	0.099
	Hot Food Holding Cabinets	4.3	39	0.075	0.95	0.047
	Fryers	0.17	1.6	0.003	0.95	0.0019
	Steamers	0.089	0.81	0.0013	0.95	0.0002
	Ice Machines	1.2	11	0.021	0.95	0.014
	Dishwashers	3.9	36	0.063	0.95	0.024
	Vending Machines	3.5	32	0.062	0.95	0.039
	Griddles	0	0	0	0.95	0
	Ovens	0	0	0	0.95	0
	Subtotal	36	330	0.63	0.76	0.44
Other	Utility Transformers	0.063	0.58	0.0011	1	0.00066
	C&I Transformers	1.1	9.9	0.019	0.77	0.015
	Residential Roofing	2.3	23	0.044	0.15	0.31
	Commercial Roofing	42	380	0.76	0.15	4.2
	Subtotal	45	420	0.82	0.15	4.6
TOTAL		1700	16000	29	0.65	26

Source: Homan et al. 2010.

As part of the ENERGY STAR program evaluation methodology, sensitivity analyses are conducted to contextualize uncertainties in data inputs and projected input parameters such as energy prices and carbon factor for electricity. Three sets of sensitivity analyses are conducted to evaluate the impact of the lower marginal carbon factor for electricity and lower ENERGY STAR sales, higher marginal carbon factor for electricity and higher ENERGY STAR sales, and lower marginal carbon factor for electricity and higher ENERGY STAR sales. The results of the sensitivity analysis are shown in Figure 37, which indicates the upper and lower bounds of the best estimate of carbon avoided and suggests that even in a “worst case” scenario, substantial carbon reductions are achieved by ENERGY STAR labeled products .

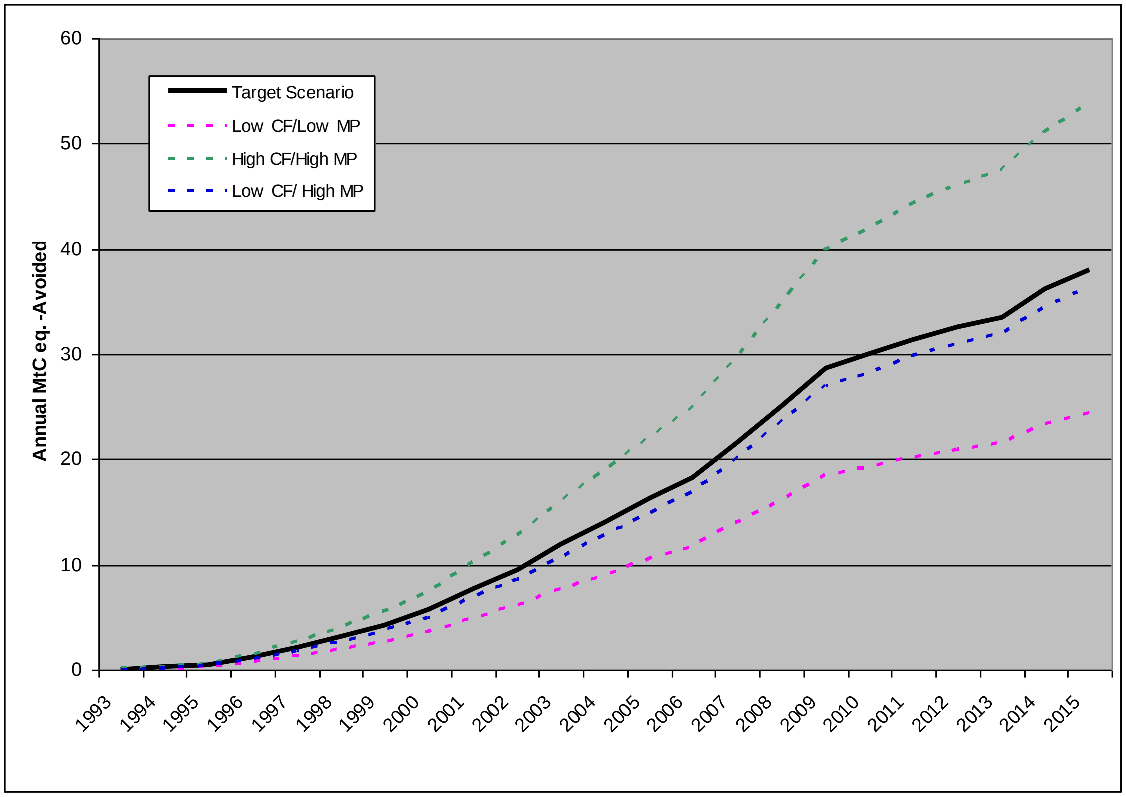


Figure 37. ENERGY STAR Evaluation Sensitivity Analysis of Carbon Savings

Source: Homan et al. 2009

3.3.2. Data Requirements and Sources

The data requirements for *ex-ante* and *ex-post* evaluation of labeling are almost identical to those identified in Section 3.2.2 for standards evaluation. The main difference, shown in red in Table 20, is that for *ex-post* evaluation, the real shipments or sales data would now be required. In the case of standards, project sales can be used in an *ex-post* evaluation, if the assumption is that all appliances sold will simply meet the basic MEPS requirement and not be of a higher efficiency than that. *Ex-post* corrections can be made to account for compliance rate. For labeling, however, data is needed on the proportion of appliances sold at the various efficiency levels associated with the categorical labeling. Whether this is a full dataset on sales or some smaller sales survey which takes a sampling of appliances sold, the data needs to be gathered in order to complete an *ex-post* impact evaluation. It could be argued that an accurate *ex-post* evaluation for standards would also require sales data by efficiency level, to account for number of units sold and unit energy savings, since actual efficiencies may not exactly match the policy levels and may indeed exceed them. There is a tradeoff between the level of accuracy sought and the related data requirements. In this case, real sales data by efficiency level can bring a higher level of accuracy than simple sales projections.

Table 20. Required and optional data requirements and sources for ex-ante/ex-post evaluation of labeling

Data type	Used in *ex-ante* or *ex-post*	Required or optional	Data source
Annual energy use per unit (UEC)	*Ex-ante, ex-post*	Required	Manufacturer test data
Existing stock	*Ex-ante, ex-post*	Required	Market data, government statistics
Market saturation (ownership, market shares)	*Ex-ante, ex-post*	Required	Market surveys
Lifetime or retirement function	*Ex-ante, ex-post*	Required	Manufacturer test data
Future shipment forecasts	*Ex-ante*	Required	Historic market data, government forecasts
UAF	*Ex-ante, ex-post*	Optional	Metered test data
NOMAD	*Ex-ante, ex-post*	Optional	Historic market data
Compliance rate	*Ex-post*	Optional	Metered or laboratory test data
Real shipments/sales	*Ex-post*	**Required**	Market data
Site-to-source energy conversion factors	*Ex-ante, ex-post*	Optional	Power plant energy data
Emission factors	*Ex-ante, ex-post*	Optional	Power plant emission data

3.3.3. Example Calculations

The following screenshots show the LEAP software's main calculations as described in section 3.3.1. Using sales data inputs, the total stock for a given year is derived by LEAP using the sum of vintage of total sales and assumed lifetime to produce the result seen in Figure 38. Using predictions for changing market shares (Figure 39) of ordinary, efficient, and highly efficient products as well as their respective UEC's (Figure 40), then the total electricity consumption can be calculated.

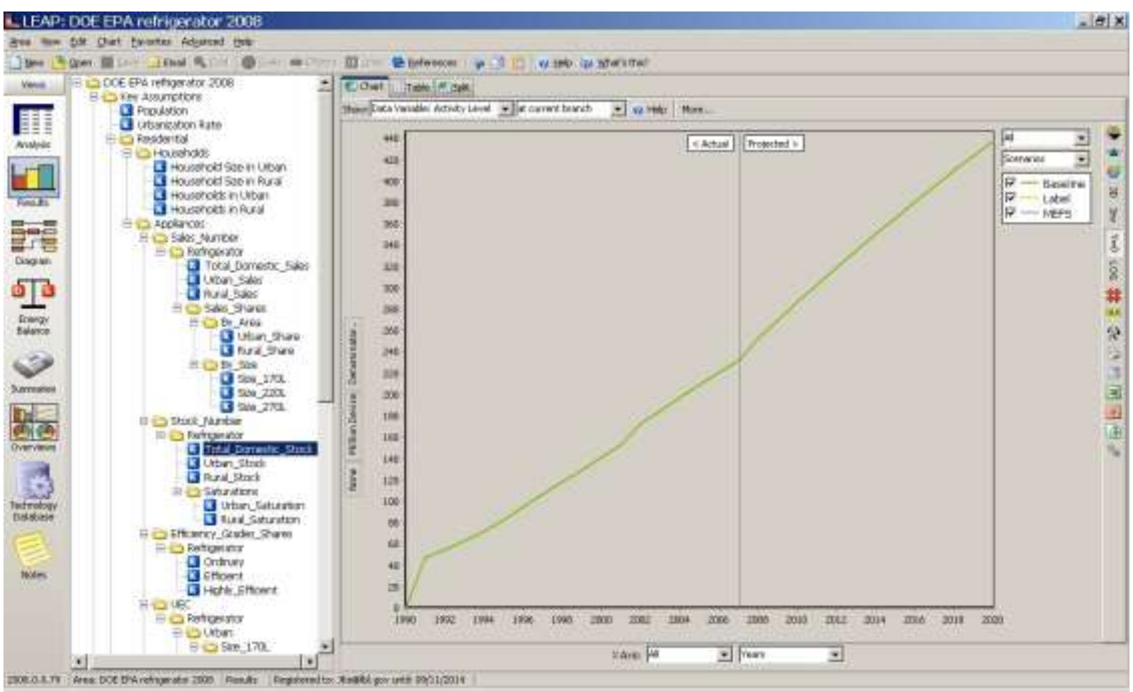


Figure 38. Growing stock of refrigerators in China, as modeled by LEAP software

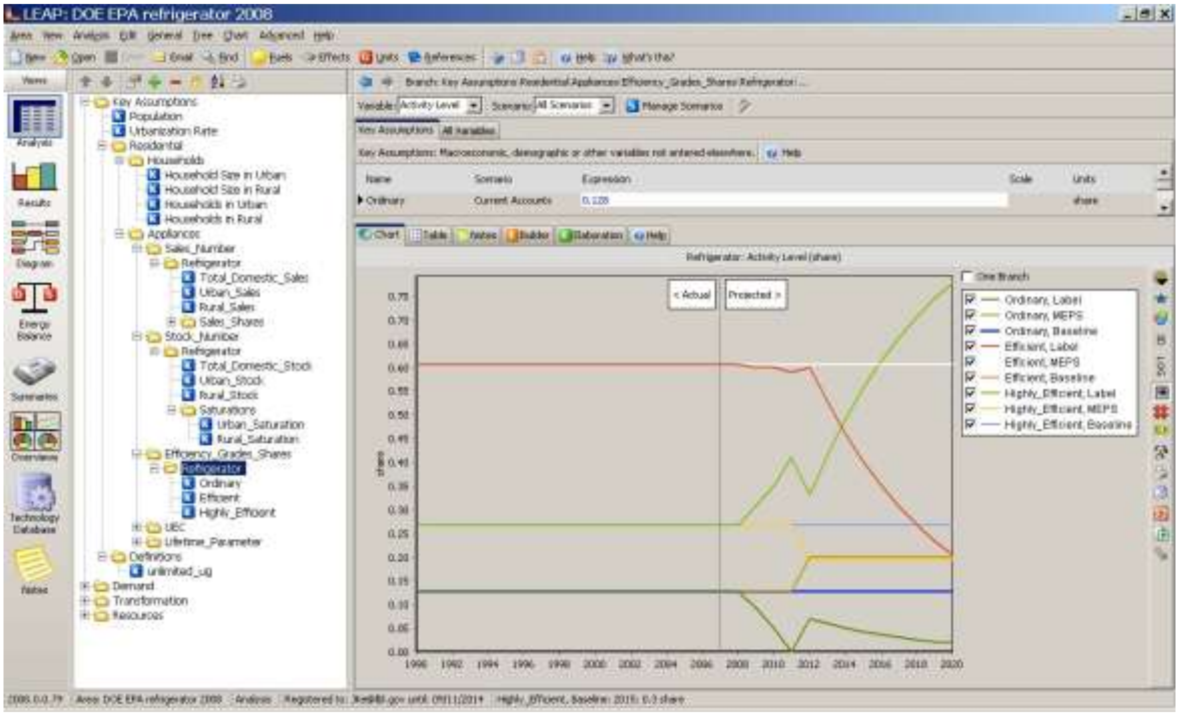


Figure 39. Changing market shares for ordinary, efficient, and highly efficient appliances in the baseline, MEPS, and labeling scenarios

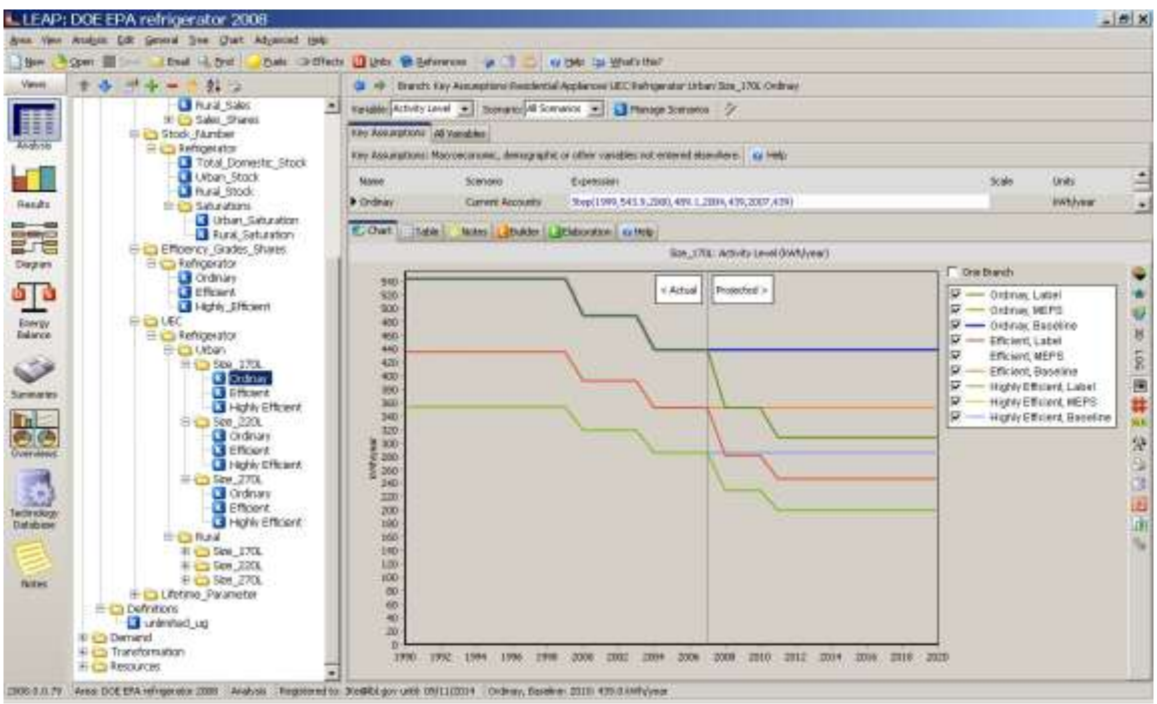


Figure 40: Changing UEC for ordinary, efficient, and highly efficient appliances

Using the three scenarios established in Table 18 in section 3.3.1 for baseline, MEPS, and labeling, the evolving market shares are set and then final electricity consumption is calculated producing the result seen below in Figure 41.

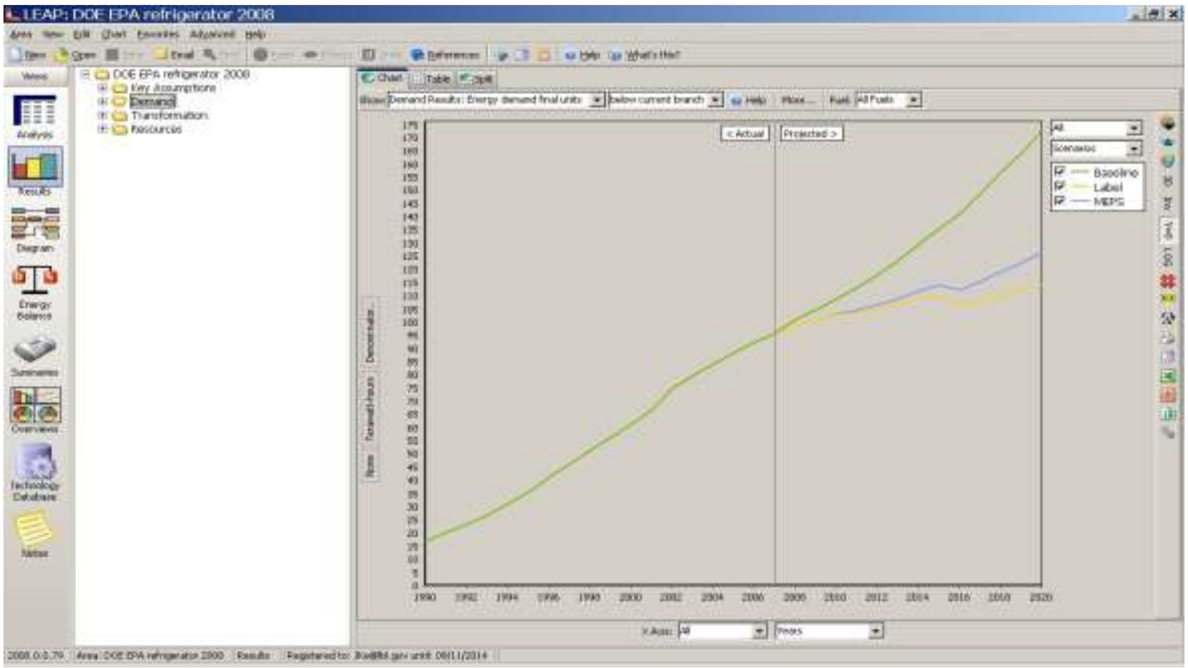


Figure 41. Resulting electricity consumption (a mirror of the result in Figure 34) based on market share MEPS and labeling scenarios described in Table 18

3.4. Impact Evaluation of Appliance Incentives Programs

Evaluation of appliance incentive programs is not as much about the application of *ex-ante* vs. *ex-post* methodologies as it is about the application of gross savings vs. net savings methodologies, particularly in the estimation of free ridership, or incentive program participants who would have adopted efficiency measures even without the incentive. This section describes incentives evaluation methodologies as well as the nuance between different measurement techniques for the data that underlies these evaluations.

3.4.1. Review of Existing Evaluation Methodologies

3.4.1.1. Ex-ante Savings Estimates

Ex-ante savings estimates (also called deemed or stipulated savings) are forecasted savings used for program and portfolio planning purposes (NEEP 2009). Engineering methods are typically used to develop ex*-ante* estimates of energy savings based on technical information from manufacturers on equipment in conjunction with assumed operating characteristics of the equipment. The two basic approaches to developing engineering estimates are engineering algorithms and engineering simulation methods. Engineering analyses are often “calibrated” with onsite data (e.g., operating hours and occupancy) to provide more credible estimates.

Engineering algorithms are typically straightforward equations showing how energy (or peak) is expected to change due to the installation of an energy efficiency measure. The accuracy of the engineering estimate depends upon the accuracy of the inputs, and the quality of data that enters an engineering algorithm can vary dramatically. Most of the savings numbers (e.g., UEC) in the preceding sections of this report are based on engineering estimates.

Engineering building simulations are computer programs that model the performance of energy-using systems in residential and commercial buildings. These models use information on building occupancy patterns, building shell, building orientation, and energy-using equipment. The input data requirements for the more complex simulation models are extensive and require detailed onsite data collection as well as building blueprints.

Ex-ante estimates are often used in *ex-post* evaluations of energy efficiency programs (after they have been "trued-up" (or adjusted) based on *ex-post* evaluations (see below)). Some states and regions in the U.S. have created databases of standardized, region-specific, *ex-ante* algorithms and associated savings estimates for conventional electric and natural gas energy efficiency measures. For example, in the U.S., there are at least 17 sources with *ex-ante* measure savings values, covering 21 states and the District of Columbia (Jayaweera et al. 2011). The databases often have an associated TRM that provides more information on the calculations and assumptions used in calculating the energy savings estimates. An example of an algorithm in a TRM is shown in Table 21.

Table 21. CFL Example from a Technical Reference Manual

Name of Measure	CFL Screw Base, Retail – Residential
Measure description	A compact fluorescent light bulb (CFL) is purchased in retail and installed in a residential location. The incremental cost of the CFL compared to an incandescent light bulb is offset via either rebate coupons or via upstream markdowns. Assumptions are based on a time of sale purchase, not as a retrofit or direct install installation.
Definition of baseline condition	The baseline is the purchase and installation of a standard incandescent light bulb.
Definition of efficient condition	The efficient condition is the purchase and installation of a compact fluorescent light bulb.
Annual energy savings algorithm	ΔkWh = ((ΔWatts) /1000) * ISR * HOURS * WHFe *Where:* *ΔWatts = Compact Fluorescent Watts (if known) * 2.95* *If Compact Fluorescent Watts is unknown use 45.7* *ISR = In Service Rate or percentage of units rebated that get installed.* *= 0.84* *HOURS = Average hours of use per year* *= 1011 (2.77 hours per day)* WHF_e *= Waste Heat Factor for Energy to account for cooling savings from efficient lighting.* *= 1.06* For example: ΔkWh = ((45.7)/1000) * 0.84 * 1011 * 1.06 = 41 kWh
Baseline adjustment	In 2012, Federal legislation stemming from the Energy Independence and Security Act of 2007 will require all general-purpose light bulbs between 40 and 100W to be approximately 30% more energy efficient than current incandescent bulbs, in essence beginning the phase out of standard incandescent bulbs. In 2012, 100W incandescents will no longer be manufactured, followed by restrictions on 75W in 2013 and 60W in 2014. The baseline for this measure will therefore become bulbs (improved incandescent or halogen) that meet the new standard. To account for these new standards, the annual savings for this measure must be reduced after 2012. For measures installed in 2010, the full savings (as calculated above in the

	Algorithm) should be claimed for the first two years, and the adjusted savings claimed for the remainder of the measure life. For measures installed in 2011, the adjustment should be made after one year's of full savings. The appropriate adjustments are provided below for 2010 and 2011 Year Installed Savings Adjustment Years of Full Savings Before Adjustment 2010 0.58 2 2011 0.50 1
Summer coincident peak kW savings algorithm	ΔkW = ((ΔWatts) /1000) * ISR * WHFd * CF *Where:* *WHFd = Waste Heat Factor for Demand to account for cooling savings from efficient lighting = 1.14* *CF = Summer Peak Coincidence Factor for measure = 0.11* For example: ΔkW = ((45.7) / 1000) * 0.84 * 1.14 * 0.11 = 0.0048 kW Note: The savings adjustment due to the shifting baseline documented above should be applied to the peak kW savings assumed in the later years.
Incremental cost	The incremental cost for this measure is assumed to be $3.
Measure life	The measure life is assumed to be 8 years.

Source: NEEP 2010

Eight states (CA, CT, MA, ME, MN, NY, TX and WI) and the Pacific Northwest region have developed a TRM, and several states are planning or considering developing TRMs (e.g., PA and IL) (Messenger et al. 2010). The deemed databases and TRMs differ in geographic coverage, but are largely regional and are administered by state regulatory commissions, state advisory committees, non-profit organizations, or utilities. Practices differ in terms of whether the use of deemed savings is mandatory for program administrators or encouraged and whether deemed values are verified, *ex-ante* or *ex-post*, by an independent party.

In their assessment of deemed savings for possible use in a national deemed savings database, Jayaweera et al. (2011) examined several possible measures, three of which are highlighted below as examples.

CFL

Residential lighting on a single fixture level is a relatively straightforward measure that is often included in a deemed database with a supporting wattage table. However, lighting is subject to many adjustment factors, depending on installation location (general house vs. living room), application (interior vs. exterior), HVAC system, and delivery mechanism (retail, direct install, socket count, etc.). Storage and removal factors also vary and must be obtained through regional studies. Additional variation arises from hours-of-use assumptions (see below). In their analysis of databases and TRMs, savings from CFLs ranged between 27 and 49 kWh.

Refrigerator

Savings are deemed, per refrigerator, based on appliance characteristics (e.g., ice through door, freezer configuration, freezer and refrigerator volume, and efficiency level). Two primary approaches are used to determine savings: (1) maximum consumption limits for baseline and efficient appliances; and (2) average region-specific appliance data. Some sources include an HVAC adjustment factor to account for refrigerators in a conditioned space interacting with the HVAC system. In their analysis of databases and TRMs, annual savings ranged between 45 and 106 kWh.

Clothes Washer

Deemed savings are tabulated by domestic water heating (DWH) and dryer fuel, and efficiency level. The clothes washer is not a standalone measure. Savings from domestic water heating and from the clothes dryer (due to less moisture in the clothes) are usually implicit. One deemed savings database calculates weighted savings values over the entire fuel mix. In their analysis of databases and TRMs, annual savings ranged between 127 and 258 kWh. Clothes washers in multifamily settings may require separate calculations to account for in-unit washers and washers in a laundry center (common area).

3.4.1.2. Evaluation of Ex-post Savings

Measurement of gross savings

Gross savings is the change in energy consumption and/or demand that results directly from program-related actions taken by participants in an efficiency program, regardless of why they participated (NEEP 2009). Typically, one compares the observed energy use of program participants with pre-project energy consumption, and then one compares this change in energy use with changes in energy use for a comparison group.

The evaluation of savings from energy efficiency programs can occur at different levels of granularity depending on the needs of program administrators and state regulatory policies:

- Average savings for one or more energy efficiency measures
- Average savings at the end use level (where more than one measure may have been installed)
- Average savings at the program level (where one or more end uses were targeted)
- Average savings at the portfolio level (where more than one program was implemented)

The type of evaluation study conducted often evolves over time. In the U.S., Messenger et al. (2010) found that jurisdictions with significant experience in implementing large-scale energy efficiency programs tend to rely on estimating savings at the measure level because these inputs are needed to assess program cost-effectiveness and the differences between planned and achieved program savings. However, in some states that are ramping up energy efficiency, evaluation efforts tend to focus on savings at the program level initially and then over time to report savings at the end use and measure level.

Methods for measuring gross energy savings

Several data analysis methods for measuring gross energy savings are available which vary in cost, precision, and uncertainty. Most monitoring and evaluation activities focus on the collection of measured data; if measured data are not collected, then one may rely on engineering calculations and *ex-ante* (deemed) savings (as described above). The most frequent types of energy efficiency programs using deemed savings are mass-market energy efficiency programs, in contrast to custom-based energy efficiency programs where unique measures are installed. Data analysis methods include basic statistical models, multivariate statistical models (including multiple regression models and conditional demand models), and integrative methods.

These methods are reflected in evaluation protocols and guidance documents, such as EVO's International Performance Measurement and Verification Protocols (IPMVP), the CPUC's California Evaluation Protocols, and NAPEE's Model Energy Efficiency Program Impact Guide.

Basic statistical models for evaluation

Statistical models that compare energy consumption before and after the installation of energy efficiency measures have been used as an evaluation method for many years. The most basic statistical models simply look at monthly billing data before and after measure installation using weather normalized consumption data (this is particularly important where weather-dependent measures are involved – e.g., heating and cooling equipment, refrigerators, etc.). If the energy savings are expected to be a reasonably large fraction of the customer's bill (e.g., 10% or more), then this change should be observable in the project's bills. Smaller changes (e.g., 4%) might also be observed in billing data, but more sophisticated billing analysis procedures are often required. This method can be used for comparing changes in energy use for program participants and a comparison group. Statistical models are most useful where many projects are being implemented. This method and the following method are rarely used in the evaluation of products that use little energy.

Multivariate statistical models for evaluation

In program evaluation, more detailed statistical models may need to be developed to better isolate the impacts of an energy-efficiency program from other factors that also influence energy use. Typically, these more detailed approaches use multivariate regression analysis as a basic tool. Regression methods are simply another way of comparing kWh or kW usage across dwelling units or facilities and comparison groups, holding other factors constant. Regression methods can help correct for problems in data collection and sampling. If the sampling procedure over- or under-represents specific types of projects among either program participants or the comparison group, the regression equations can capture these differences through explanatory variables. Two commonly applied regression methods are conditional demand analysis and statistically adjusted engineering models.

End-use metering

Energy savings can be measured for specific equipment for specific end uses through end-use metering. This type of metering is conducted before and after a retrofit to characterize the performance of the equipment under a variety of load conditions. The data are often standardized for variations in operations, weather, etc. End-use metering reduces measurement error (assuming the metering equipment is reliable) and reduces the number of control variables required in models. In addition, the meter can calculate the energy change on an individual piece of equipment in isolation from the other end-use loads.

Short-term monitoring

Short-term monitoring refers to data collection conducted to measure specific physical or energy consumption characteristics either instantaneously or over a short time period. This type of monitoring is conducted to support evaluation activities such as engineering studies, building simulation and statistical analyses. Examples of the type of monitoring that can take place are spot watt measurements of efficiency measures, run-time measurements of lights or motors, temperature measurements, or demand monitoring. Short-term monitoring is gaining increasing attention as evaluators realize that for certain energy efficiency measures with relatively stable and predictable operating characteristics (e.g., commercial lighting and some motor applications), short-term measurements will produce gains in accuracy nearly equivalent to that of longer-term metering at a fraction of the cost. To illustrate the type of activities that are conducted in a monitoring study, an example of a monitoring study research plan for refrigerators as shown below:

1). Monitor the hourly electricity usage and room temperatures for 160 existing refrigerators and 30 new Energy Star replacement units in a sample of homes drawn from four target programs;
2). Use the program implementers to screen potential sites, perform site data collection about the refrigerators and households, and deploy metering equipment as part of their regular work;
3). Deploy the metering in five waves spread over the course of year to reflect varying weather and other seasonal effects;
4). Retrieve the meters within two to three weeks of deployment;
5). Develop a model of indoor temperatures by analyzing the temperature data with weather data and information about refrigerator location and occupant-reported thermostat settings;
6). Analyze each site's usage and temperature data to develop an estimate of annual usage, correcting for differences in temperature between the metering period and an estimate of the site's annual temperature;
7). Assess the accuracy of the different program implementers' refrigerator auditing techniques including adjusted rated usage and short-term metering;
8). Attempt to develop an improved refrigerator auditing technique based on refrigerator and site characteristics directly observable during a typical field audit, such as rated usage, refrigerator age and condition, household size, and estimated indoor temperatures., and assess the value of this approach compared to short-term (<=2 hour) metering;
9). Assess the energy usage of new replacement refrigerators compared to their rated usage values
10). Develop program savings adjustment factors, to the extent feasible, based on the program implementer-estimated energy savings and actual usage results from the detailed data
11). Develop load shape estimates for the existing and new refrigerators to assess load impacts (Blasnik 2004).

Integrative methods

Integrative methods combine one or more of the above methods to create an even stronger analytical tool. These approaches are rapidly becoming the state of the practice in the evaluation field. The most common integrative approach is to combine engineering and statistical models where the outputs of engineering models are used as inputs to statistical models. These methods are often called Statistically Adjusted Engineering (SAE) methods or Engineering Calibration Approaches (ECA). Although they can provide more accurate results, integrative methods typically increase the complexity and expense. To reduce these costs while maintaining a high level of accuracy, a related set of procedures has been developed to leverage high cost data with less expensive data. These leveraging approaches typically utilize a statistical estimation approach termed ratio estimation that allows data sets on different sample sizes to be leveraged to produce estimates of impacts.

Best methods

There is no one approach that is 'best' in all circumstances (either for all program types, evaluation issues, or all stages of a particular program). The costs of alternative approaches will vary and the selection of evaluation methods should take into account program characteristics and the kind of load and schedule for the load before the retrofit. The load can be constant, variable, or variable but predictable, and the schedule can either be known (timed on/off schedule) or unknown/variable. The monitoring approach can be selected according to the type of load and schedule.

Adjustments to deemed savings

Standard evaluation practice strongly recommends that when using deemed savings (see above) it is crucial to verify a sample of installations to ensure that the measures were actually installed and working per the specifications defined for using the deemed savings value. In some cases, depending on the measure and application (in residential or commercial sector), hours of use are measured (since this is such a key assumption used in the calculation of deemed savings). In the U.S., eight states have or are considering an audit requirement to verify a sample of installations resulting from efficiency programs (Messenger et al. 2010).

Net savings

Net savings is the total change in energy consumption and/or demand that is attributable to an energy efficiency program. This change may include, implicitly or explicitly, the effects of free drivers, free riders, energy efficiency standards, changes in the level of energy service, and other causes of changes in energy consumption or demand (NEEP 2009). Evaluation measurement methods are well documented and relatively standardized for determining gross energy savings for energy efficiency programs. However, there is much less agreement on the value and methods that should be used to estimate net savings. In contrast to the parameters used to adjust gross savings, the net savings parameters cannot be directly measured because they are at least partially based on a counterfactual – what would have happened without the program (intervention) – not what actually did happen. This is the reason that net savings estimates can be controversial. One of the most important concepts to understand within a technical measurement approach is that net savings is a behavior metric that adjusts gross savings to account for how a program influences the decision-making processes of the participants or people in the marketplace. Thus, net savings evaluation approaches measure changes in decision behavior, and the evaluation approach must document how the program changed end users' decision behaviors. Another key issue is how to assess the broader "net" market effects of energy efficiency programs.

The concept of net energy savings is fairly simple: what were the true effects produced by a program or intervention in terms of energy savings, separated out from what would have otherwise occurred absent the program or intervention? Unfortunately, this simple concept is exceptionally difficult to measure in practice, particularly in a way that meets specific reliability standards for accuracy or comparability. This problem is compounded in current practice in the U.S. because there are two general conditions that impact the ability to estimate net impacts. These are the questions of definition and technical measurement.

The definition of what constitutes net energy impacts can be state-specific, in some cases program-specific, requiring the measurement approach to be tailored to meet the applicable definition for a specific regulatory jurisdiction. The difference in definitions can have a substantial impact on the estimate, as well as on the evaluation method that is used. For example, in California (2004-2009), net energy savings are defined by the California Public Utilities Commission to be gross energy savings minus the energy savings from free riders. In this case, the gross energy savings are reduced to account for what a specific evaluation methodology can identify as a program-induced installation, subtracting out savings from instillations that are driven by other factors. The following formula represents the current California definition:

Net savings = gross savings – free riders

On the other hand, in New York, net energy savings are defined by the New York Public Service Commission as gross energy savings, minus savings from free riders, plus energy savings due to participant spillover and market effects. Participant spillover is the savings from program participants who, as a result of the program, installed additional energy efficiency measures, but who did not obtain a program incentive for those additional measures. Market effects are the market level savings that resulted from program influences on the market and the operations of that market (sometimes referred to as nonparticipant spillover, since these end users did not participate in the program and did not obtain a program incentive for those measures), but the market for energy efficiency was affected by the program. The following formula represents the New York definition:

Net savings = gross savings –free riders + participant spillover + market effects.

In some states, market effects are not equivalent to nonparticipant spillover, since program participants as well as nonparticipants are affected by market effects. For example, in Wisconsin, depending on the program, the evaluation of net savings may focus either on: (1) free riders only, (2) free riders and participant spillover only, (3) free riders, participant spillover, and nonparticipant spillover, or (4) total market-level net impacts, without any effort to disaggregate by spillover type.

Because the market effects of a program can be as large as or larger than the program's gross savings, the resulting quantification of net effects from one state to another can be very different for the same program, rebating the same measures, targeting the same customers. The definitional difference alone makes comparing a net effect from one program to the next problematic, particularly if the evaluation approach varies from state to state. Similarly, in a carbon-focused world, the definition of net effects can result in large and significant differences in reported carbon reductions resulting from the same program operating in two different jurisdictions.

Once the definitional issue is addressed, typically through a regulatory decision establishing the definition of net savings for a specific state, the technical issues associated with measurement must be addressed. The measurement of net energy savings can be accomplished using a variety of different approaches.

Free riders

Free ridership can be evaluated either explicitly or implicitly. The most common method of developing explicit estimates of free ridership is to ask participants what they would have done in the absence of the project (also referred to as "but for the project" discussions). Based on answers to carefully designed survey questions, participants are classified as free riders (yes or no) or assigned a free ridership score. Project free ridership is then estimated as the proportion of participants who are classed as free riders. Two key problems arise in using this approach: (1) very inaccurate levels of free ridership may be estimated, due to questionnaire wording and unreliable self-reports; and (2) there is no estimate of the level of inaccuracy, for adjusting confidence levels.

Another method of developing explicit estimates of free ridership is to use discrete choice models to estimate the effect of the program on customers' tendency to implement measures. The discrete choice is the customer's yes/no decision whether to implement a measure. The discrete choice model is estimated to determine the effect of various characteristics, including project participation, on the tendency to implement the measures.

A method for calculating implicit estimates of free ridership is to develop an estimate of savings using billing analysis (as described above) that may capture this effect, but does not isolate it from other impacts. Rather than taking simple differences between participants and a comparison group, however, regression models are used to control for factors that contribute to differences between the two groups (assuming that customers who choose to participate in projects are different from those who do not participate). The savings determined from the regression represent the savings associated with participation, over and above the change that would be expected for these customers due to other factors, including free ridership.

Spillover

Spillover refers to the reductions in energy consumption and/or demand caused by the presence of an energy efficiency program, beyond the program-related gross savings of the participants and without financial or technical assistance from the program (NEEP 2009). There can be participant and/or nonparticipant spillover. Participant spillover is the additional energy savings that occur when a program participant independently installs energy efficiency measures or applies energy saving practices after having participated in the efficiency program. Non-participant spillover refers to energy savings that occur when a program nonparticipant installs energy efficiency measures or applies energy savings practices as a result of a program's influence.

Estimates of spillover are determined using one of the following approaches, similar to those used in the evaluation of free riders (NAPEE 2007):

a. Self-reporting surveys in which information is reported by participants and non-participants without independent verification or review.
b. Enhanced self-reporting surveys in which self-reporting surveys are combined with interviews and documentation review and analysis.
c. Statistical models that compare participants' and non-participants' energy and demand patterns, their knowledge about efficiency options, and/or the trade-offs they are willing to make between efficiency options and the costs of purchasing and installing them.
d. Stipulated net-to-gross ratios (ratios that are multiplied by the gross savings to obtain an estimate of net savings) that are based on historic studies of similar programs.

Market Effects

Market effects evaluations estimate a program's influence on encouraging future energy efficiency projects because of changes in the marketplace. The evaluation focuses on the changes in the structure or functioning of a market, or the behavior of participants in a market, that results from one of more program efforts (NEEP 2009; Vine 2012). Typically, the resultant market or behavior change leads to an increase in the adoption of energy-efficient products, services, or practices. While all categories of programs can be assessed using market effects evaluations, they are primarily associated with market transformation programs that indirectly achieve impacts and resource acquisition programs that are intended to have long-term effects on the marketplace. In the U.S., only a few market effects evaluations have been conducted (Vine 2012). An example of the type of indicators that are examined in a market effects study from an air-conditioning program is shown in Table 22.

Table 22. NYSERDA's Keep Cool Program Indicators

Program Outputs	Short-Term Outcomes	Intermediate-Term Outcomes	Long-Term Outcomes
Number of air conditioners surrendered Number and dollar value of bounties and other incentives paid Number of units demanufactured and amount of material diverted from the waste stream Number of ads placed, impressions, and ad value Number of contacts with consumers at events	Change in awareness of NYSERDA program and ENERGY STAR(R) (consumers, multi-family building owners, relevant small commercial owners, and retailers) Effectiveness of TV advertising versus other advertising venues Knowledge and ability of retail staff to promote ENERGY STAR(R) RACs and efficient TTW units Stocking of ENERGY STAR(R)	Perceived benefits of ENERGY STAR(R) product purchases Degree of subsequent ENERGY STAR(R) product purchases given past ENERGY STAR(R) experience Frequency and content of communication to others concerning experience with ENERGY STAR(R) product Retailers indicate that ENERGY STAR(R) RACs and efficient TTW units are	In conjunction with other ENERGY STAR(R) efforts: - Eliminated barriers - Reduced waste by recycling of old units - Increasing market share and penetration - Sustained change in market behavior - Persistent energy savings - Emissions reductions

Number of retailers active in the program Number of facilities for collecting old room air conditioners (RAC) and through-the-wall (TTW) units. Number of calls to hotline Number of website hits	RACs and efficient TTW units Increased availability and use of turn-in facilities Immediate Peak Reduction and KW and KWh savings resulting from program activities (within program sales and outside of program sales as induced by the program) Immediate peak reduction resulting from consumer behavioral change (load shifting) due to the energy tips marketing campaign Benefit-cost ratios (if environmental benefits of demanufacturing can be properly included)	profitable to them as evidenced by their stocking patterns Retailers indicate (by actions or words) that promoting and selling ENERGY STAR(R) RACs and efficient TTW units are a profitable activity KW and KWh savings	

Source: NYSERDA 2004

The following types of approaches are used to collect data in market effects evaluations (Vine 2012):

a. Review of program material and related literature
b. Review of program administrator's program data
c. Review of baseline sales and market data (e.g., building practice and code compliance)
d. Telephone surveys and in-person interviews with customers, retailers, distributors, manufacturers, contractors, consultants, builders, government officials, program managers and evaluators
e. In-home (onsite) audits
f. Stocking inventories

The following types of analysis are conducted in market effects evaluations (Vine 2012):

a. Statistical models that compare participants' and non-participants' knowledge about efficiency options, purchase behavior, and/or the trade-offs they are willing to make between efficiency options and the costs of purchasing and installing them.
b. Multivariate regression modeling for analyzing awareness, availability, and the program's effect on pricing of specific energy efficiency products, by controlling for other factors that impact sales of energy-efficient measures, including income, education, housing characteristics, and utility rates.
c. Delphi or expert panel approach, in which gross savings and penetration of technologies and practices are estimated and presented to panel members, who are then asked to attribute savings to energy efficiency programs and other factors; it is essential that there be at least two rounds of Delphi surveys, with the first round results summarized and presented in the second

round survey so panel members can understand and learn from each other in developing the final attribution estimates.

3.4.2. Data Requirements and Sources

Data requirements for an impact evaluation of incentive programs will depend on the scope and budget of the evaluation, as shown in Table 23. For instance, deemed values pulled from TRM's for similar studies will be a much less expensive evaluation option than performing original data collection and statistical analysis.

Table 23. Required and optional data requirements and sources for impact evaluation of incentive programs

Data type	Required or optional for gross energy savings	Required or optional for net energy savings	Data source
Annual energy savings per unit product or per building	Required	Required	Deemed values, IPMVP approach, or statistical analysis
Number of participants and non-participants	Required	Required	Surveys
Normalizing factors (HDD, CDD)	Required	Required	Weather station
Free riders	Optional	Required	Surveys, econometric methods, deemed value
Participant spillover	Optional	Required	Surveys, econometric methods, deemed value
Market effects (participant & nonparticipant spillover)	Optional	Required	Surveys, econometric methods & market analysis
Site-to-source energy conversion factors	Optional	Optional	Power plant energy data
Emission factors	Optional	Optional	Power plant emission data

If one is interested in only gross energy savings, then the first three data types are required. If one is interested in net energy savings, then the first six data types are required.

As the types of incentives for efficiency measures can vary widely, so can their associated evaluation methods. While some evaluation methods can be entirely prescriptive based on deemed savings estimates, other evaluation methods are custom based on site-specific conditions and employ end-use metering. While it is more common to see deemed savings values used for evaluations of retail CFL and refrigerator rebate programs, it is more common to see end-use metering and other forms of monitoring for those energy efficiency measures that are less common, as shown in Figure 42. In between, surveys are often used as a way of modifying deemed savings values or corroborating monitoring results.

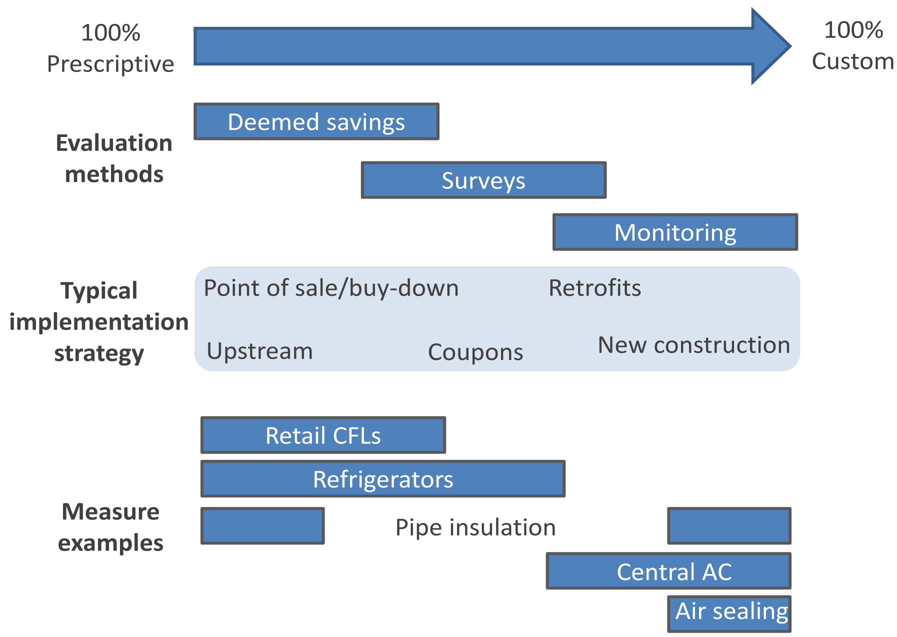


Figure 42. Prescriptive vs. custom evaluation methods and related implementation strategies and measures

Source: Adapted from Dent and Enterline 2012

3.4.3. Example Calculations

This section will provide important examples of incentives evaluation from the U.S. to illustrate how gross and net savings estimates are calculated, including survey examples for figures such as free ridership. A study conducted for the state of Maine's energy efficiency program offering CFL discounts and coupons began with deemed savings estimates, which were then modified based on data gathered from over 400 telephone surveys. The following equations were used to calculation gross savings per unit and total demand savings:

$$Gross\ Savings\ =\ ([UEC_b - UEC_e] * \frac{Hours\ of\ use}{day} * \frac{Days}{year} * In\ service\ Rate)\ /\ \frac{1000\ Watts}{kW}$$

$$Demand\ Savings\ =\ Number\ of\ measures * ([UEC_b - UEC_e] *\ In\ service\ Rate)\ /\ \frac{1000\ Watts}{kW}$$

From this equation, the variables include the change in UEC, hours of use, and in-service rate (how many CFL's were being used of those purchased). Efficiency Maine at first assumed an in-service rate of 100% and 2.7 hours per day of usage. After conducting telephone surveys, the numbers changed significantly, as seen in Table 24. For example, the in-service rate was actually just 60-72%. The hours of use ranged from 2.3 to 4.8 hours.

Table 24. Key study parameters from Efficiency Maine CFL discount and coupon program evaluation

	CFLs discount	2003-2005 coupon CFLs	2006 coupon CFLs
Volume of Products	199,336	283,591	545,192
Wattage reduction (W)	45	45	45
Daily hours of use (hours)	4.8	2.3	3.2
In-service Rate	60%	72%	66%
Assumed lifetime (years)	4.6	9.5	6.8
Gross Annual Energy Savings per Unit (kWh)	47	27	35
Gross Energy Savings (MWh)	43,375	73,279	128,605
Freeridership rate	29%	20%	20%
Spillover rate	23%	46%	30%
Net-to-Gross ratio (1 + SO – FR)	0.94	1.26	1.10

Source: NMR 2007a

Demand savings were calculated to be 22.1 MW using the in-service rates from the surveys. Most efficiency studies seek to know how much of this demand savings is coincident with peak demand (e.g., are CFLs operating when demand is at its peak?). Efficiency Maine recognizes winter weekday hours of 5-7 pm as its winter peak period. Installed program CFLs were turned on an average of 33.6% of the time during these hours, according to telephone surveys, thus decreasing the demand savings amount to a peak incident demand savings value of 7.5 kW.

The next step involved estimating the free ridership and spillover rates, so that a net savings amount could be determined from the gross savings.

$$Net\ Savings\ =\ Gross\ Savings\ *\ (1\ +\ Spillover\ rate - Free\ ridership\ rate)$$

Free ridership was defined by the study as those purchases that would have been made by the participants on their own within three months without the provided incentive. In the telephone surveys, the evaluators obtained the following information:

- Awareness of efficient lighting product prior to program purchase
- Intention to buy the product at the same time or within three months of the program purchase
- Willingness to pay average retail price ($5/CFL) for a specific number of products purchased

Similarly, spillover purchases were defined as products purchased since the program purchase and products that were purchased without any coupons but affected by the program. The respondents were asked if they had been influenced by the experience of the program purchase to make the additional purchases (NMR 2007a).

From these telephone surveys, free ridership rates were estimated at 20-29% while spillover rates were estimated at 23-46%. Given that the effects of these two factors offset each other to some degree, the net-to-gross ratio ended up being either very close to or well above 1. Efficiency Maine compared the values it found from its telephone surveys to those found in similar program evaluation studies in the northeastern U.S., shown in Table 25.

Table 25. Comparison of key values for Efficiency Maine CFL program evaluation with other similar regional program evaluations

	In-service rate	Wattage reduction rate	Daily hours of use	Freeridership rate	Spillover Rate
Current study findings	66%	45.0	3.2	20%	30%
2004 MA/RI/VT study	62%	48.7	2.7	11%	22%
2002-2003 NH RLP study	62%	40.9	4.7	19%	4%
2000-2001 NU SLC/RL study	70%	52.0	3.4	-	-
1998 Starlights study	73%	54.8	3.4	-	-

Source: NMR 2007a

A 2004 study by the New York State Energy Research and Development Authority (NYSERDA) on rebates offered for ENERGY STAR room air-conditioners (AC) used a unique tracking method to verify the efficiencies of AC units purchased and AC units retired. In addition to rebates offered on new units sold when old units were turned in, NYSERDA also contracted two companies to pick-up and recycle old AC units that were turned in, and retailers were paid storage incentives to cover storage fees for offering to provide the turn-in facility and storage. The incentives were $75 a unit initially, then lowered to $35 a unit because of the decreasing price of ENERGY STAR room AC units as compared to non-ENERGY STAR room AC units. The storage incentives offered were $15-25 per unit. Since all turned-in units were being collected at common points and retail data were also being tracked, NYSERDA could verify the model and efficiency of each unit retired and each unit purchased. Annual operating hours were necessary to calculate the energy savings and were determined based on the number of cooling degree days in the region (NYSERDA 2004).

A 2007 study on a Massachusetts rebate program for ENERGY STAR clothes washers detailed how free ridership was evaluated. In the program, rebates were given for CEE Tier 2 and 3 products (see Table 26 for more detail on product efficiency).

Table 26: Summary of Clothes Washer Efficiency Specifications for Massachusetts ENERGY STAR rebate program

Year	Federal Standard	ENERGY STAR	CEE Tier 1	CEE Tier 2	CEE Tier 3
2006	>1.04 MEF	>1.42 MEF	>1.42 MEF, <9.5 WF	>1.60 MEF, <8.5 WF	>1.80 MEF, <7.5 WF (A), <5.5 WF (B)
2007	>1.26 MEF	>1.72 MEF, <8.0 WF	>1.80 MEF, <7.5 WF	>2.00 MEF, <6.0 WF	>2.2 MEF, <4.5 WF

Source: NMR 2007b; Note: MEF = Modified Energy Factor; WF = Water Factor

A telephone survey was administered to a random selection of program participants (those who had purchased an ENERGY STAR clothes washer and received a rebate). Initial questions asked the participant to identify how familiar they were with ENERGY STAR, while the free ridership questions asked the participant to identify how likely they would have been to purchase the model without incentives. The full list of questions is listed below in

Table 27.

Table 27. Survey to determine free ridership in ENERGY STAR clothes washer program evaluation

ES1. Are you familiar with the ENERGY STAR label on household products? The label is a blue and white label with the word "energy" followed by a five-pointed star. ENERGY STAR labels are used by the Environmental Protection Agency—the EPA—and the Department of Energy to identify and label highly energy-saving appliances and other products for consumers. Before this description, how familiar were you with the ENERGY STAR label? Would you say you were:

1. Very familiar
2. Somewhat familiar
3. Slightly familiar, or
4. Not at all familiar before being read this description? **[Skip to #FR1]**
5. (Don't know) **[Skip to #FR1]**

ES4. Would you say that all ENERGY STAR-qualified clothes washers are pretty much equally energy efficient, or are some ENERGY STAR-qualified models significantly more energy efficient than others?

1. All are pretty much equally energy efficient
2. Some are significantly more energy efficient than others
3. (Don't know)

FR1. If you had not received the $100 rebate from [COMPANY], how likely would you have been to purchase the same clothes washer at full retail price? Use a scale from 0 to 10, where 0 is "definitely would have chosen a different model," and 10 is "definitely would have chosen the same model even without the rebate." [11=DON'T KNOW]

FR2. [If #FR1 < 7 AND #ES1 ≤3] If you had not received the $100 rebate from [COMPANY], how likely would you have been to purchase an ENERGY STAR qualified clothes washer model? Use a scale from 0 to 10, where 0 is "definitely would have chosen a different model," and 10 is "definitely would have chosen the same model even without the rebate." [11=DON'T KNOW]

FR3. [If #FR2 >6 AND #ES4 =2] If you had not received the $100 rebate, would you have purchased a minimally energy efficient ENERGY STAR qualified model, a moderately energy efficient model, or a highly energy efficient model?

1. Minimally energy efficient model
2. Moderately energy efficient model
3. Highly energy efficient model
4. (Don't know)

Source: NMR 2007b

Using the responses gathered from the survey, the evaluator devised the following algorithm to determine level of free ridership from non-free rider to full-free rider. The algorithm is described in Table 28. For instance, a full-free rider is a participant who responded a definite response (7-10 on a scale of 10) to the first free ridership question asking directly whether the participant would have purchased the model without the incentive.

Table 28. Free ridership algorithm for CEE Tier 3 purchasers

Level of Free rider	Question FR1	Question FR2	Question FR3
Non Free rider	Responses 0-6 and respondent is not aware of ENERGY STAR label to #ES1	All responses 0-6	None
Partial: CEE Tier 1	n/a	Responses 7-10 if respondent is not aware of different levels of energy efficiency to #ES4	Response 1
Partial: CEE Tier 2	n/a	n/a	Response 2

Partial: CEE Tier 3	n/a	n/a	Response 3
Full Free Rider	Responses 7-10	n/a	n/a

Source: NMR 2007b

In this study, the instances of free ridership were found to be very high at 78%. The results for the CEE Tier 3 rebates can be seen in Table 29.

Table 29. Self-Reported Free Ridership Estimates

	Tier 3 rebates	Percent
Non free rider	3,425	12%
Partial free rider – CEE Tier 1/regular ES (2006—MEF 1.42)	571	2%
Partial free rider – CEE Tier 2 (2006—MEF 1.6)	285	1%
Partial free rider – CEE Tier 3 (2006—MEF 1.8)	571	2%
Full free rider	22,822	78%
Unknown		5%
Total	27,674	100%

Source: NMR 2007b;

Note: red square is coordinated with red square in Table 30 to indicate same values

Using a survival function similar to that used in the U.S. stock model described in Section 3.3.1, the evaluators determined the average lifetime of clothes washers to be 14.6 years as shown in Figure 43.

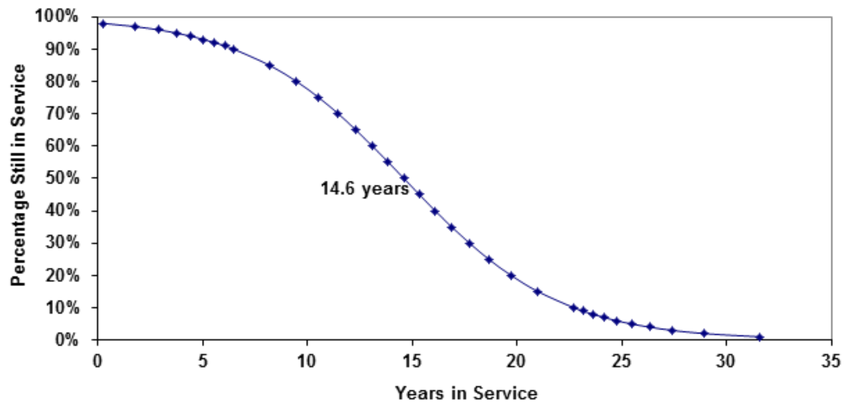


Figure 43. Survival function for clothes washers

The survival function helps determine which of those ENERGY STAR units that were purchased in the year of the rebate will be still in use at a later date. Then, based on the free ridership and UEC_e values, the total lifetime savings of the program can be calculated as shown in Table 30

Table 30: Lifetime Electricity Savings Estimates Based on Self-Reported Free Ridership

		number of units Tier 3				savings MWh Tier 3				
Year	% in use	non-freerider	partial FR 1	partial FR 2	partial FR 3	non-freerider	partial FR 1	partial FR 2	partial FR 3	Total Tier 3 Savings
2006	98%	3,425	571	285	571	1,306	117	41	49	1,513
2007	97%	3,399	567	283	567	1,296	116	40	49	1,502
2008	96%	3,366	561	280	561	1,284	115	40	49	1,487
2009	95%	3,323	554	277	554	1,267	113	40	48	1,468
2010	93%	3,268	545	272	545	1,246	111	39	47	1,444
2011	91%	3,199	533	267	533	1,220	109	38	46	1,413
2012	89%	3,113	519	259	519	1,187	106	37	45	1,376
2013	86%	3,008	501	251	501	1,147	103	36	43	1,329
2014	82%	2,879	480	240	480	1,098	98	34	42	1,272
2015	78%	2,726	454	227	454	1,040	93	32	39	1,204
2016	73%	2,549	425	212	425	972	87	30	37	1,126
2017	67%	2,349	391	196	391	896	80	28	34	1,038
2018	61%	2,130	355	177	355	812	73	25	31	941
2019	54%	1,897	316	158	316	724	65	23	27	838
2020	47%	1,660	277	138	277	633	57	20	24	733
2021	41%	1,426	238	119	238	544	49	17	21	630
2022	34%	1,204	201	100	201	459	41	14	17	532
2023	28%	999	167	83	167	381	34	12	14	441
2024	23%	817	136	68	136	312	28	10	12	361
2025	19%	660	110	55	110	252	22	8	10	292
2026	15%	527	88	44	88	201	18	6	8	233
2027	12%	417	69	35	69	159	14	5	6	184
2028	9%	327	54	27	54	125	11	4	5	144
2029	7%	255	43	21	43	97	9	3	4	113
2030	6%	198	33	16	33	75	7	2	3	87
2031	4%	153	25	13	25	58	5	2	2	68
2032	3%	118	20	10	20	45	4	1	2	52
2033	3%	91	15	8	15	35	3	1	1	40
Lifetime MWh savings										21,861

Source: NMR 2007a

Note: red square is coordinated with red square in Table 29 to indicate same starting values while % in use is pulled from survival function.

Methodologies for evaluating incentive programs vary widely, and evaluators use many unique approaches to determine gross and net savings values. In general, surveying is used in most incentive evaluations as a way to increase the accuracy of gross and net savings estimates to reflect true program impacts. The balance between cost of the surveying and evaluation and the accuracy of the evaluation should be weighed carefully depending on the type and scope of measure.

3.5. From Energy Savings to Carbon Savings

A full description of methodologies for calculating carbon savings from gross or net energy savings is beyond the scope of this paper. In general, a number of approaches can be used, including:

- Average carbon multiplier approach
- Hourly weighted average carbon multiplier approach
- Hourly dispatch carbon emissions calculation approach

There are a number of uncertainties that exist in converting energy savings into carbon savings, however, including:

- What fuel type was saved
- Efficiencies of the generation facilities impacted
- The hour the savings occur over 8,760 h/year
- Generation mix for any given hour over the effective useful life of the savings
- How generation facilities are cycled or how to accurately predict unit cycling and the relationship to demand and savings
- Effective useful life of the savings projections

4. Conclusions

Most evaluation handbooks' first recommendation is to get quality data and to start collecting it as early as possible. This report has highlighted a wide set of international methodologies and studies for this report, showing the array of evaluation options available. The basic requirement for standards and labeling impact evaluations is a high quality dataset for appliance sales in order to build a working stock model. A number of options exist for improving the accuracy of evaluation estimates, such as calculating compliance rates and correction factors using surveys, metering data, or laboratory test data. In measuring the change in unit energy consumption from a base model to a more efficient model, deemed savings values can be used and verified or modified after doing some amount of metering in the field if the budget allows. For evaluation of incentives, many evaluators calculate net to gross ratios using a variety of survey techniques in order to account for participant behavior and market interactions. In general, most evaluators agree on a basic set of methodologies for appliance energy efficiency program evaluation, but there is a wide variation in techniques used to gather the required and optional data which has an associated range of costs.

Acknowledgments

We would like to express our deep appreciation to the China Sustainable Energy Program of the Energy Foundation and the Collaborative Labeling and Appliance Standard Program (CLASP) for providing funding for this project. We would especially like to thank Kevin Mo and Steven Zeng. We would also like to thank Steve Kromer and CLASP for their helpful review of this report.

This work was supported by the China Sustainable Energy Program of the Energy Foundation and the Collaborative Labeling and Appliance Standard Program (CLASP) through the U.S. Department of Energy under Contract No. DE-AC02-05CH11231.

References

Blasnik, M., 2004, “Measurement & Verification of Residential Refrigerator Energy Use: Final Report, 2003 – 2004 Metering Study.”

Broc, J. et al, 2009, “The development process for harmonized bottom-up evaluation methods of energy savings,” Wuppertal, Germany: EMEEES.

Brown, M., Nevius, M., 2007, “Evaluation Methods for Achieving Energy Efficiency Policy Objectives: Metering the Unmetered Resource,” Presentation to IPU Advanced Regulatory Studies Program.

[CNIS] China National Institute of Standardization, 2012, “White Paper Energy Efficiency Status of Energy-Using Products in China (2012),” Beijing, China: Standards Press of China.

[CPUC] California Public Utilities Commission, 2006, “California Energy Efficiency Evaluation Protocols: Technical, Methodological and Reporting Requirements for Evaluation Professionals.” San Francisco, CA: California Public Utilities Commission.

CPUC, 2008, “California Long Term Energy Efficiency Strategic Plan.” San Francisco, CA: California Public Utilities Commission.

Dent, S., Enterline, S., 2012, “Energy Efficiency Programs: Deemed Savings & Technical Reference Manuals,” EPA Energy Efficiency EM&V Webinar Series.

[DOE] Department of Energy, 2011a, “Technical Support Document in support of Energy Conservation Standards for Residential Refrigerators, Refrigerator-Freezers, and Freezers.” Washington, DC: Department of Energy.

DOE, 2011b, “Residential Refrigerators, Refrigerator-Freezers and Freezers Final Rule Analytical Tools,” Washington, DC: Department of Energy.

Ecofys et al, 2006, “Guidelines for the monitoring, evaluation and design of energy efficiency policies - How policy theory can guide monitoring & evaluation efforts and support the design of SMART policies,” Utrecht, Netherlands: European Directive on Energy Efficiency (AID-EE).

[EVO] Efficiency Valuation Organization, 2012, “International Performance Measurement and Verification Protocol: Concepts and Options for Determining Energy and Water Savings Volume 1.”

[EE Strategies] Energy Efficient Strategies, 2002, “Evaluation of the Australian energy efficiency standards and labeling program: Report to the Australian Greenhouse Office and the National Appliance and Equipment Energy Efficiency Committee,” Canberra, Australia: Australia’s Department of Climate Change and Energy Efficiency.

EE Strategies, 2010, “Evaluation of energy efficiency policy measures for household refrigeration in Australia,” Canberra, Australia: Australia’s Department of Climate Change and Energy Efficiency.

Energy Market Innovations, 2011, “Consumer Electronics Television Initiative Market Progress Evaluation Report,” Prepared for Northwest Energy Efficiency Alliance.

[EPA] Environmental Protection Agency, 2011, “National Awareness of ENERGY STAR® for 2010: Analysis of 2010 CEE Household Survey”, Washington, DC: EPA Office of Air and Radiation, Climate Protection Partnerships Division.

Greenblatt, J., Hopkins, A., Letschert, V., Blasnik, M., 2012, “Energy use of US residential refrigerators and freezers: function derivation based on household and climate characteristics,” *Energy Efficiency* (2012): DOI 10.1007/s12053-012-9158-6.

Homan, G., Sanchez, M., Brown, R., 2010, "Calendar Year 2009 Program Benefits for ENERGY STAR Labeled Products." Berkeley, CA: Lawrence Berkeley National Laboratory.

Itron, 2011, "DEER Database: 2011 Update Documentation," California Public Utilities Commission.

Jayaweera, T., H. Haeri, A. Lee, S. Bergen, C. Khan, A. Velonis, C. Gurin, M. Visser, A. Grant, and A. Buckman. 2011. Scoping Study to Evaluate Feasibility of National Databases for EM&V Documents and Measure Savings. Berkeley, CA: Lawrence Berkeley National Laboratory.

KEMA Inc., 2005, "CREST: The California Residential Efficiency Saturation Tool," http://calresest.kemainc.com/

Kushler, M., Nowak, S., White, P., 2012, "A National Survey of State Policies and Practices for the Evaluation of Ratepayer-Funded Energy Efficiency Programs." Washington, DC: American Council for an Energy Efficient Economy.

Lapillonne, B., D. Bosseboeuf, and S. Thomas, 2009, Top-down Evaluation Methods of Energy Savings: A Summary Report, Grenoble, Switzerland: Enerdata; Paris, France: ADEME; and Wuppertal, Germany: Wuppertal Institute.

Larsen, TF, Petersson, K., Naeraa, R., 2012, "Estimation Tool for National Effects of MEPS and Energy Labeling," Rome, Italy: International Energy Program Evaluation Conference, June 2012.

Larsonneur, P., et al, 2009, "EMEES bottom-up case application 5: Energy-efficient cold appliances and washing machines," Wuppertal, Germany: EMEES.

Luttmer, M., et al, 2006, "Evaluation of labeling of appliances in the Netherlands," Utrecht, Netherlands: European Directive on Energy Efficiency (AID-EE).

McNeil, M. et al, 2008, "Global potential of energy efficiency standards and labeling programs," Berkeley, CA: Lawrence Berkeley National Laboratory, November 2008, LBNL-760E.

McNeil, M. et al, 2011, "Business Case for Energy Efficiency in Support of Climate Change Mitigation, Economic and Societal Benefits in the United States," Berkeley, CA: Lawrence Berkeley National Laboratory, June 2011, LBNL 4682E.

McNeil, M., Letschert, V., de la Rue du Can, S., Ke, J., 2012, "Bottom-Up Energy Analysis System – Methodology and Results," Washington, DC: Collaborative Labeling and Appliance Standards Program.

Messenger, M., R. Bharvirkar, B. Golemboski, C. Goldman, and S. Schiller. 2010. "Review of Evaluation, Measurement and Verification Approaches Used to Estimate the Load Impacts and Effectiveness of Energy Efficiency Programs." Berkeley, CA: Lawrence Berkeley National Laboratory.

Meyers, S. et al, 2011, "Energy and Economic Impacts of U.S. Federal Energy and Water Conservation Standards Adopted from 1987 through 2010," Berkeley, CA: Lawrence Berkeley National Laboratory, December 2011, LBNL-5291E.

Meyers, S., McMahon, J., Atkinson, B., 2008, "Realized and Projected Impacts of U.S. Energy Efficiency Standards for Residential and Commercial Appliances," Berkeley, CA: Lawrence Berkeley National Laboratory, March 2008, LBNL-63017.

[NAPEE] National Action Plan for Energy Efficiency, 2007, "Model Energy Efficiency Program Impact Evaluation Guide," Washington, DC: U.S. Environmental Protection Agency and Department of Energy.

Natural Resources Canada, 2012, "Improving Energy Performance in Canada." Ottawa, Canada, Natural Resources Canada.

[NMR] Nexus Market Research, 2007a, "Process and Impact Evaluation of the Efficiency Maine Lighting Program," Cambridge, MA: Nexus Market Research.

NMR, 2007b, "Estimates of Net Impact of the 2006 Massachusetts ENERGY STAR Appliances Program, Clothes Washer component," Cambridge, MA: Nexus Market Research.

[NYSERDA] New York State Energy Research and Development Authority, 2004, "New York Energy $mart Program Evaluation and Status Report: Final Report, Volume 2," Albany, NY: NYSERDA.

[NEEP] Northeast Energy Efficiency Partnerships, 2009, "Glossary of Terms (Version 1.0)," Lexington, MA: NEEP.

NEEP, 2010. Mid-Atlantic Technical Reference Manual, Version 1.1.

Quantec, 2007, "Statewide Codes and Standards Market Adoption and Noncompliance Rates," Portland, OR: Quantec

Research Into Action et al, 2010, "Research Supporting an Update of BPA's Measurement and Verification Protocols."

RLW Analytics, 2008, "Coincidence Factor Study Residential Room Air Conditioners," Lexington, MA: NEEP.

Schiller, S., C. Goldman and E. Galawish. 2011. "National Energy Efficiency Evaluation, Measurement and Verification (EM&V) Standard: Scoping Study of Issues and Implementation Requirements," Berkeley, CA: Lawrence Berkeley National Laboratory, LBNL-4265E.

[SoCal Edison] Southern California Edison, 2009, "SCE's 2009-2011 Energy Efficiency Program Plan Implementation Plans."

SRC International, NOVEM, et al, 2001, "A European ex-post evaluation guidebook for DSM and EE service programs," SAVE project No. XVII/4.1031/P/99-028.

TecMarket Works, 2004, "The California evaluation framework," San Francisco, CA: California Public Utilities Commission.

Ting, M., Rufo, M., 2010, "Learning from the past and predicting the future: Linking program evaluations to energy efficiency planning studies," Counting on Energy Programs: It's Why Evaluation Matters, Paris, France: International Energy Program Evaluation Conference, June 2010.

Van Buskirk, R., 2012, "An Adoption Curve Fitting Method for Estimating Market Efficiency Improvement and Acceleration (working paper)," Washington, DC: U.S. Department of Energy.

Vermont Energy Investment Corporation (VEIC), 2010, "State of Ohio Energy Efficiency Technical Reference Manual,"

Vine, E. 2012. "Transforming the Energy Efficiency Market in California: Key Findings, Lessons Learned and Future Directions from California's Market Effects Studies," in the *Proceedings of the 2012 ACEEE Summer Study of Energy Efficiency in Buildings*. Washington, DC: American Council for an Energy Efficient Economy.

Vine, E. et al, 2012, "Emerging issues in the evaluation of energy-efficiency programs: the US experience," *Energy Efficiency* (2012) 5:5–17.

Vine, E., Sathaye, J., 1999, "Guidelines for the Monitoring, Evaluation, Reporting, Verification, and Certification of Energy-Efficiency Projects for Climate Change Mitigation."

Vine, E., C.H. Rhee, and K. Lee, 2006, "Measurement and evaluation of energy efficiency programs: California and South Korea," *Energy* 31: 1100-1113.

Vreuls, H. et al, 2009, "General bottom-up data collection, monitoring, and calculation methods," Wuppertal, Germany: EMEES.

Vreuls, H., 2005, "Evaluating energy efficiency policy measures and DSM programs: Volume 1, Evaluation Guidebook," Paris, France: International Energy Agency Implementing Agreement on Demand-Side Management Technologies and Programs (IEA-DSM).

Waide, P. 1997. "Refrigerators: Developments in the European Market." Energy Efficiency in Household Appliances. P. Bertoldi, A. Ricci, and B. Wajer, eds. Springer.

Wiel, S., 2002. "Energy Efficiency Experiences: Standards, Labels and Other Energy Efficiency Policies." Presentation at the APEC.

Wiel, S., McMahon, J., 2005, "Energy Efficiency Labels and Standards: A Guidebook for Appliances, Equipment, and Lighting, 2nd edition," Washington, DC: Collaborative Labeling and Appliance Standards Program (CLASP).

Wuppertal Institute, 2009, "Measuring and reporting energy savings for the Energy Services Directive – how it can be done," Wuppertal, Germany: EMEES.